

US should back research cloning

The US Senate is expected to vote on legislation that would ban cloning by the end of this month. Supporting cloning for research is both right and in the national interests of the United States.

As technology continues to advance, countries must decide how they will deal with the issue of human cloning for reproduction or research. So far, several nations have placed strong restrictions on therapeutic cloning; others are moving towards such restrictions, and a few have staked out positions in favour of therapeutic cloning. After months of acrimonious debate, the United States must now decide what it will do.

All legislators can agree that it would be wrong now to make a walking, talking, real-life human clone. The National Academy of Sciences also supports that position. But its Institute of Medicine has rightly said that its objections to the safety of reproductive cloning do not apply to research cloning. Indeed, some scientists say that research cloning could yield stem cells that could be used to grow therapeutic tissues for patients with diseases such as Parkinson's. They also say that studying stem cells made from the cells of diseased patients could help us understand why people with the same genetic make-up get sick or stay well.

Opponents of research cloning say there is no proof that it will yield any cures. They also say that adult stem cells are more promising and less controversial. They have gained Congressional and public support by tapping into widespread fears about biotechnology, which some worry is careening quickly down a slippery slope towards the commodification of the human species. But such fears do not represent a sensible basis for a ban on research cloning, which is likely to give insights into the processes that underlie a host of debilitating diseases (see *Nature* **414**, 567; 2001).

The Senate is now moving towards a showdown on this issue. Two bills have been introduced. Senator Sam Brownback (Republican,

Kansas) introduced a bill that would ban cloning for any purpose. His rivals, led by Senator Dianne Feinstein, (Democrat, California), have introduced competing legislation that would allow scientists to clone embryos for research. And senators eager to air their views on the issue are calling for a vote on the matter in the next few weeks. Brownback is said to have nearly 50 supporters, but for technical reasons a bill is unlikely to be passed unless 60 senators support it.

Advocates of therapeutic cloning have outlined scenarios that would make the Senate more likely to pass a bill that would allow research cloning, such as amending the Brownback bill to allow research. In this way, senators could save face by simultaneously voting for Brownback and for research.

However, any bill that does pass the Senate must be reconciled with the House bill in a conference. The Brownback bill is virtually identical to a House cloning ban that was passed last July. So it would speed through the conference committee. But Senate and House negotiators are unlikely to compromise if the Senate votes to allow therapeutic cloning. So the result of this month's Senate debate is likely to be either that President Bush signs a bill that bans cloning for any purpose, or that he does not sign any cloning bill at all. The issue could also spill over into the appropriations process this autumn, when senators try to force rules through the Congress by attaching them to necessary spending bills.

The Congress has strongly supported the National Institutes of Health in recent years because it wants the United States to be a world leader in biomedical research. The Senate should continue its strong support of biomedical science, and act in the national interest, by refusing to pass a ban on research cloning. ■

Peer review reviewed

A controversial change to the peer-review process of Germany's principal funding agency is long overdue.

The Deutsche Forschungsgemeinschaft (DFG), Germany's main funding agency, is symbiotically embedded in the German scientific community. Its referees, who decide which research to fund, are elected every four years in a process led by scientific societies and involving all tenured researchers.

This system means that researchers by and large have confidence in 'their' agency. But it has not kept pace with the development of science. The DFG senate therefore agreed this week on a reform, which — if approved by the general assembly in July — will strengthen the initiative of the DFG's programme managers and bring additional expertise to the reviewing process.

Predictably, there is friction. Some interpret the move as a turn away from democratic ideals, or a presumptuous attempt to increase the power of the DFG's bureaucratic machinery. But such responses are misplaced. The 650 elected referees cannot cope with the growing complexity of science, nor with the 20,000 or so grant applications every year. Reviewing procedures take too long, interdisciplinarity gets short shrift, and unconventional or risky projects have little chance of funding.

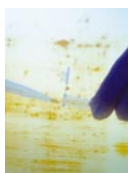
In recent years the DFG has called in a growing number of additional experts, both from Germany and from abroad, to support the reviewing process. In future, if the new proposals are approved, this will become the rule. The DFG's programme managers, who are respected scientists themselves, work cheek-by-jowl with the scientific community and tend to be good at overseeing developments in fast-moving fields. The DFG's elected referees will still have the final say, and will be able to scrutinize the administrators' choice of reviewers.

In reviewing its processes, the DFG also hopes to create an additional bottom-up stimulus to its funding priorities. The reform should bring the agency closer to common international standards. It will hardly revolutionize the German science system, but it is perhaps a long-awaited sign that the system is becoming more dynamic, and more accepting of new scientific ideas. The quality of science, and Germany's attractiveness to foreign scientists, can only benefit. The DFG has shown that it is willing to shift. Its general assembly and the community at large should welcome the changes. ■





Full mouse
Draft sequence of
mouse genome roars
into public view
p106



Big vs small
Proteomics leaves
NIH with large-scale
debate
p107



Done deal
Plea bargain clears
Japanese biologist
accused of espionage
p108



Badger baiting
Controversial
tuberculosis study
resumes
p109

Academy protests as agriculture agency abandons visa renewals

Virginia Gewin, Washington

The US Department of Agriculture (USDA) is washing its hands of visa renewals for researchers at its facilities who are not US citizens.

The action — which will disrupt the careers of up to 200 postdoctoral researchers and visiting scientists — has alarmed US scientific leaders. They fear that it could presage wider difficulties for foreign researchers in the country in the wake of last September's terrorist attacks.

Bruce Alberts, president of the National Academy of Sciences, brought the USDA's new policy to the attention of John Marburger, President George Bush's science adviser, during an open meeting of the academy's members in Washington at the end of last month.

"It's not a reasonable policy," Alberts said in an interview. "It's going to hurt us more than help us." But he expressed confidence that the policy will be modified or reversed now that Marburger has been made aware of it.

The USDA — citing a lack of ability to monitor the visa-renewal process with the care that now is expected of government agencies — says that it will no longer sponsor any visa applications or extensions.

Among those directly affected are 200 postdocs, visiting scientists and temporary employees at the agency's Agricultural

Research Service laboratories, a network of 100 laboratories that employs about 2,000 scientists. Those affected can continue working until their current visas expire.

The USDA's policy change was triggered by its concerns about another group of immigrants — foreign physicians seeking J-1 visa waivers at the end of their training. In the past three years, the USDA has recommended over 300 of these waivers, which exempt physicians from the usual requirement of returning to their native country after US medical training, provided that they work in rural areas that are short of doctors.

"It is now important for us to know more about the medical doctors and scientists we sponsor," says USDA spokeswoman Alisa Harrison. She adds that the policy remains under review and that the USDA will do what it can to continue to work with the researchers — cooperating, for example, with universities that might sponsor their visa renewals.

The agricultural board of the National Academy has asked Alberts to complain formally to the government about the USDA's new policy. After he complained informally to Marburger on 29 April, Marburger thanked him for raising the issue and pledged to look into it. But a week later, a spokeswoman for Marburger declined to specify what he would do to address the concern, and



Seeds of doubt: the USDA's new policy could leave foreign researchers unable to renew their visas.

referred questions on the policy to the USDA.

Visas are not the only security issue worrying US science leaders. In his address to the National Academy meeting, Alberts listed several threats that face US science, and placed "the possibility of excessive restrictions on scientific publication, motivated by security concerns", at the top of the list. ■

S. BAUER/USDA

Radiologist's confirmation ends NIH power vacuum

Erika Check, Washington

The US Senate has confirmed radiologist Elias Zerhouni as director of the National Institutes of Health (NIH), filling a 28-month void at the helm of the world's largest biomedical research agency.

Zerhouni, who was born and medically trained in Algeria, will arrive at the NIH from the Johns Hopkins University School of Medicine in Baltimore, Maryland, where he was vice-dean for research. The Senate confirmed Zerhouni on 3 May, just three days after a smooth confirmation hearing at which he stressed the importance of multidisciplinary science, the translation of

basic research into medicine, and better access to new technologies for researchers.

He tactfully told senators that he would support whatever decision the Congress comes up with on the issue of therapeutic cloning (see page 103). He said that he backed President Bush's decision to use federal funds for research only on stem-cell lines that already exist, adding that if researchers needed more of them, he would "be the first to assemble that information" for consideration by policy-makers.

Zerhouni also got firmly behind the creation of the Institute of Biomedical Imaging and Bioengineering at the NIH —

a development that was instigated by Congress but opposed by Harold Varmus, the NIH's previous permanent director.

A healthy backlog of work awaits Zerhouni at his new office in Bethesda, Maryland. Since Varmus left, the NIH has lost several institute directors, and Tommy Thompson, the health secretary, has moved aggressively to centralize some of the agency's functions at his own office in Washington DC. "The most important role of a director right now is to re-establish morale and momentum," Zerhouni told the senators, "and to make the agency even more effective than it has been." ■

Draft mouse genome makes public debut

Erika Check, Washington

A draft assembly of the mouse genome has been made public by the Mouse Genome Sequencing Consortium.

The achievement is unlikely to attract the same public attention as last year's release of the human genome — but researchers say it could be just as important to science and medicine.

"The mouse genome turns a spotlight on all sorts of features that were not apparent from the human genome," says Eric Lander, director of the Center for Genome Research at the Whitehead Institute in Cambridge, Massachusetts, one of three sequencing

centres involved in the project. "It just didn't make sense to wait until the paper came out to make sure people knew about it."

The sequence — posted on 6 May on GenBank, the main public archive of genomic information, and two other websites — covers an estimated 96% of the genome, with each base sequenced an average of seven times.

Direct comparisons between the mouse and human genomes will yield valuable information, Lander says, including the identification of regulatory regions and even new genes. But the genome will also serve as a powerful tool for biomedical researchers, who are planning to breed mice with certain

genotypes and watch how they develop.

"Many laboratories have been labouring to identify mouse variants that associate with particular phenotypes," explains Francis Collins, director of the National Human Genome Research Institute in Bethesda, Maryland. "This can now be done with a few clicks of a mouse and we can move on to understanding the biology and trying to apply this to good purpose."

Collins and Lander played down the role of the public consortium's rivalry with Celera, the biotechnology company in Rockville, Maryland, in the timing of their announcement. Last April, Celera said it had assembled a draft sequence of the mouse genome that included sixfold coverage of each base, but made it available only to subscribers to its own database.

Celera is now said to be planning publication of its mouse genome in some shape or form — perhaps releasing a detailed analysis of a single chromosome. "We do have a plan to do some publishing, but we can't talk about it in any detail at this time," says Celera spokesman Robert Bennett.

♦ www.ncbi.nlm.nih.gov



Of mice and men: comparisons between the human and mouse genomes might unearth new genes.

Climate panel to focus on regions

Jim Giles, London

Rajendra Pachauri, the new chair of the Intergovernmental Panel on Climate Change (IPCC), has pledged that under his stewardship the global advisory group will place more emphasis on regional assessments of climate change, and on its socio-economic impact.

Pachauri, the director of the Tata Energy Research Institute in New Delhi, says that the prospect of accelerated ice melting on the Himalayas is the type of regional question that the panel needs to address.

"Five hundred million people depend on these glaciers for their water supply," says the 62-year-old energy economist. "What kind of response do we have? To take an extreme example, we could move highly water-intensive industry out of that region. If you have 20 to 30 years to do that you are not going to cause major disruption."

Pachauri assumed his new position on 28 April after an unusual, contested election in which he defeated the incumbent chair, atmospheric scientist Robert Watson.

In an interview, Pachauri sought to draw a line under the election, in which many climate-change researchers expressed their support for Watson. In an article written just after the election, Al Gore, the former US

presidential candidate, branded Pachauri as the "let's drag our feet candidate". Pachauri says that he had regarded Gore as a friend. "It's unfortunate that he should have played politics in this manner," he says.

Pachauri adds that the failure of the Bush administration to renominate Watson — which led to the latter being nominated late, by other nations — could be attributed to Watson's links to the previous administration of Bill Clinton. "When a new president takes over, everyone gets changed — including the janitor," he says.

The new IPCC chair says that he is now busy setting the agenda for the panel's fourth assessment report, which is due in 2007. He also wants to produce a report to give input to negotiations that begin in 2005 to set greenhouse-gas emission targets for 2013 onwards under the Kyoto Protocol. Such a report, he says, might include an assessment of technology issues, such as the potential of carbon sequestration.

Watson, meanwhile, has immersed himself in a new project. He is launching a \$500,000 feasibility study on an advisory mechanism to inform governments and international bodies on the scientific issues facing global agriculture, including the implications of genetically modified crops. ■

Welfare amendment omits lab animals

Virginia Gewin, Washington

In a victory for biomedical research lobbyists, rodents and birds are to remain outside the scope of the Animal Welfare Act, the main US animal-protection law.

Under an amendment to the current farm bill placed by Senator Jesse Helms (Republican, North Carolina), the US Department of Agriculture will be forbidden from extending the act to cover the animals most widely used in research. The farm bill, with the amendment intact, was approved by a conference between the House and the Senate on 26 April, and President Bush has promised to sign it after both houses vote for the agreed measure.

The act oversees the housing, health-care and transport conditions for animals used in research. Universities and biomedical research lobbyists had campaigned energetically for its continued exclusion of mice, rats and birds.

But animal-rights groups say that the success of the amendment calls into question the relevance of the entire act, as rats and mice alone account for 95% of the animals used in laboratory research in the United States. ■

NIH ponders issues of scale in protein push

Erika Check, Washington

The US National Institutes of Health (NIH) faces a tough balancing act between funding large institutes and supporting individual laboratories in the still-emerging discipline of proteomics, said scientists at a workshop on the topic.

Twenty-four senior researchers were invited to the NIH campus in Bethesda, Maryland, late last month to tell its managers how best to invest in proteomics, which seeks to establish the structure and function of all human proteins. But the limitless scope of such an enterprise clearly perplexed some of the participants, many of whom voiced suspicion of any large-scale approach akin to the Human Genome Project.

In theory, proteomics could encompass the study of all known human proteins, complexes of proteins, and interactions of proteins with small molecules. But unravelling such a vast array of information in a systematic way would require new and expensive technologies. And although some researchers at the workshop argued that the NIH should invest in large centres that would develop and exploit such technologies, most said that they



Drip-feed or deluge? Researchers are at odds over the best way to generate data on human proteins.

favour approaches that would encourage individual investigators to study small parts of large, complex protein networks.

Several delegates voiced concern that a high-throughput approach to finding proteins would be hugely expensive but yield little insight. John Yates, a mass-spectrometry specialist at the Scripps Research Institute in La Jolla, California, dismissed some such efforts as “low-IQ proteomics”.

Other researchers argued that without uniform standards, such as the ones that have been adopted for gene microarrays, even existing data on proteins will be difficult to replicate between laboratories. Still others said that large data sets will prove useless without individual investigators who are motivated to wade through them to gain new insights into their favourite systems.

“We need a benevolent champion for individual molecules,” said Steve Clarke, a biochemist at the University of California, Los Angeles (UCLA). “We have tremendous amounts of information already and so much of it goes unused.”

But others declared support for a ‘big-science’ approach to the subject. Ruedi Aebersold, a co-founder of the Institute for Systems Biology in Seattle, Washington, joined Peer Bork of the European Molecular Biology Laboratory, Heidelberg, in advocating large-scale protein profiling. They said that the NIH should set up pilot centres stocked with mass spectrometers that would profile the proteins expressed in the bodies of healthy and sick patients.

Other participants questioned whether individual researchers will be able to use the data generated by high-throughput technologies. “How are we going to connect all these data sets to the laboratories of the people asking the questions?” asked William Studier, a structural-genomics expert at the Brookhaven National Laboratory in New York state.

David Eisenberg, director of the US Department of Energy’s Laboratory of Structural Biology and Molecular Medicine at UCLA, said: “I think there was consensus that it’s probably not going to be possible to have a scaled-up human proteomics project in the next couple of years.”

NIH managers attending the workshop included Francis Collins, director of the National Human Genome Research Institute, who said he would use its guidance to help frame his institute’s next five-year plan, which is due out next spring.

Phoenix rises to genomic challenge

Rex Dalton, San Diego

A major facility that will perform large-scale gene-expression analysis on tissue samples and release its findings to the public is set to be established in Phoenix, Arizona.

The board of the International Genomics Consortium — a broad group of pharmaceutical companies, universities and other backers that will operate the facility (see *Nature* 410, 855; 2001) — is expected to announce within the next month that the centre will be built in Phoenix and directed by Jeffrey Trent, currently scientific director of the National Human Genome Research Institute, part of the National Institutes of Health (NIH), at Bethesda, Maryland.

The consortium has already attracted more than \$42.5 million from pharmaceutical companies and foundations to advance its first genomic project on cancer. And discussions already are under way to expand the work to other diseases, including Alzheimer’s, multiple sclerosis, diabetes and heart and lung disorders.

In the past year, 19 major US medical centres have agreed to provide the consortium with samples of malignant and normal tissue for gene-expression analysis. Genetic data would then be available for use by any scientist worldwide — with no intellectual property restrictions attached. The concept

will be similar to that of the SNP Consortium, a non-profit, public-private group set up in 1999 to provide researchers with single-nucleotide polymorphisms, the common variations in human DNA.

At the genomic consortium, microarray analysis will be done on DNA samples from tumours and normal tissue. The resultant expression records of the messenger RNA turned on in the tissue samples will be made available to researchers on the consortium’s open database.

The consortium was the idea of Richard Mallery, a Phoenix-based lawyer who chairs its board, and has won support from others who want to expand research and biotechnology in Arizona. The state of Arizona and other local backers are putting together an incentive package worth tens of millions of dollars to help support the facility. The state legislature is expected to pass the necessary financing measures this month.

Other locations, including Atlanta and Baltimore, were considered for the project but Phoenix was always top of the list. Trent, the consortium’s chief scientific adviser, has wanted to return to his home state to lead the venture. He says he hasn’t yet given official notice that he will leave the NIH, but adds that this could happen shortly.

■ www.intgen.org

Psychiatrist offered visiting post after Toronto drugs row

David Spurgeon, Montreal

David Healy, the University of Wales psychiatrist who sued the University of Toronto, its dean of medicine and the head of one of its teaching hospitals in a dispute over academic freedom, has won an appointment as a visiting professor at the university's medical school.

Healy, who has written and lectured widely on psychological drugs, had been offered the job of clinical director of the mood and anxiety disorders programme at the university's affiliated Centre for Addiction and Mental Health (CAMH).

But the offer — which included a university professorship — was withdrawn after Healy gave a lecture at the university that was critical of 'psychotropic' drugs, including Prozac, made by Eli Lilly, which contributes funds to the CAMH (see *Nature* 413, 240; 2001).

The withdrawal resulted in a letter to the university president from 27 eminent scientists calling it "an affront to the standards of free speech and academic freedom". Healy's Can\$9.4-million (US\$6-million) lawsuit charged that the university denied him academic freedom and that he was defamed as a scientist and physician during the university's attempts to justify the job withdrawal.

The CAMH website carries a joint statement from Healy, itself and the university, saying that they are "pleased to announce the settlement of all litigation and other outstanding disputes".

"Although Dr Healy believes that his clinical appointment was rescinded because of his November 2000 speech at CAMH, Dr Healy accepts assurances that pharmaceutical companies played no role" in the reversal, the statement says.

The announcement adds that Healy will continue to write and speak on issues concerning pharmaceutical companies, research and academic freedom. Healy will visit the university for one week each year during the next three years.

Vic Catano, president of the Canadian Association of University Teachers, calls the settlement "a complete vindication for Dr Healy". ■

♦ www.camh.net

♦ www.caut.ca

Spying charges dropped as Japanese biologist cuts deal

David Cyranoski, Tokyo

One of the two Japanese biologists facing unprecedented charges of espionage in the United States has struck a plea bargain and pleaded guilty to a less serious felony.

Hiroaki Serizawa, a molecular biologist currently at the University of Kansas, entered his plea — saying that he was guilty of making false statements to investigators — as part of a deal with prosecutors reached at the Northern District of Ohio Court on 1 May.

The deal was struck less than two weeks before Serizawa was due to go on trial accused of helping Takashi Okamoto steal DNA samples and cell lines from the Cleveland Clinic Foundation, where Okamoto worked, and transport them to Japan. Okamoto took up a post at the Institute of Physical and Chemical Research (RIKEN) in Japan in 1999.

US prosecutors have now formally asked the Japanese government to extradite Okamoto, and Serizawa appears to be ready to testify against him.

Serizawa and Okamoto were charged in May 2001 with violating the 1996 Economic Espionage Act, transporting stolen property and conspiracy (see *Nature* 411, 225–226; 2001). The severity of the charges caused a furore in Japan, where their actions were widely viewed as consistent with Japanese researchers' comparatively relaxed approach to the transfer of research materials.

The charge that Serizawa has admitted — which is still classed as a felony — refers to statements he made about his relationship with Okamoto when he was approached by the Federal Bureau of Investigation in September 1999. As part of his bargain, Serizawa has agreed to cooperate in the case against Okamoto, who still lives in Japan, although he resigned from his position with RIKEN last summer after the charges were brought.

Serizawa now faces a fine of up to \$5,000 and possibly a prison term of not more than six



Plea bargain: Hiroaki Serizawa faces reduced charges and can stay in the United States.

months. A conviction on the original charges could have put him in jail for up to 15 years.

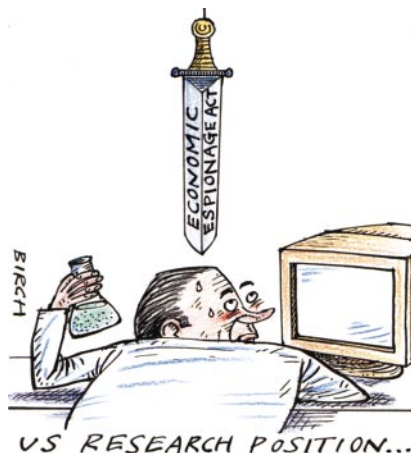
"The plea agreement recognizes that Serizawa did not conspire with, aid or abet Dr Okamoto in the theft of trade secrets or research materials," says John McCaffrey, one of Serizawa's lawyers. The deal allows Serizawa to stay and work in the United States, something McCaffrey says is "unprecedented" for a foreign national with a felony conviction.

But Serizawa is deeply in debt as a result of his legal bills, and his supporters say that the episode could damage him as he seeks a tenured research position. "He will have an uphill battle," says David Zapol, a researcher at Pharsight, a biotechnology company in Mountain View, California, who set up a support group for Serizawa. "He didn't deserve any of this."

US officials would not give further details of the plea bargain. "All I can say is that the original charges against Okamoto still stand and that Serizawa has agreed to cooperate in that case," says Christian Stickman, the assistant US Attorney handling the case in Ohio.

The case marks the first application of a clause in the Economic Espionage Act that expressly prohibits the stealing of trade secrets for the benefit of a foreign entity, and which some say could damage science.

"People may be wary of communicating research results as well as passing materials to colleagues," claims Joan Conaway, a molecular biologist at the Stowers Institute for Medical Research in Kansas City who supervised Serizawa's postdoctoral research. "That could have a devastating effect on research." ■



Neurologists strike gold in drug screen effort

Alison Abbott

A number of drug leads and the prospect of a deeper understanding of neurodegenerative diseases have emerged from the first 'community screen' in the United States.

The programme, run by the National Institute of Neurological Disorders and Stroke (NINDS) in Bethesda, Maryland, brought together a wide community of scientists in a search for interesting compounds. Between them, they tested about 1,000 drugs, most of which are already in use as treatments for conditions unrelated to neurological illness.

The drugs, whose identities were kept secret from the researchers, were screened over a six-month period in 29 different laboratory tests that model neurodegenerative disorders such as Alzheimer's disease and amyotrophic lateral sclerosis (ALS).

The investigators involved met in Washington last month to discuss the results. "It was very exciting," says Jeffrey Rothstein, a neuroscientist at Johns Hopkins Medical School in Baltimore. "We had all cranked away blind for six months, and then we learnt that there had been hits in different assays with classes of drug used clinically for indications which have nothing to do with the nervous system."

Allan Tobin, director of the Brain Research Institute at the University of California, Los Angeles, also participated in the experiments. "This is the most exciting enterprise I have been involved in since we discov-

ered the Huntington's gene," he says, referring to the 1993 gene find that he organized.

The screening effort is part of wider push by the National Institutes of Health to improve the translation of basic research results into clinical use. The tests were done by investigators who already have grants either from NINDS or from three voluntary organizations that co-sponsored the programme — the ALS Association, the Huntington's Disease Society and the Hereditary Disease Foundation.

Three-quarters of the drugs tested are already approved by the US Food and Drug Administration (FDA), and the rest were selected because they influence brain activity. These included controlled substances, such as cannabinoids, the active ingredients in marijuana.

Investigators are most excited about what the screens, in which so many different parameters were altered by exposure to so many different classes of drug, will reveal about the biology of neurodegenerative diseases, rather than the prospect of immediate clinical benefits. Some say that this is analogous to the sort of information that emerges from DNA microarray studies that measure expression of hundreds of thousands of genes, or from large-scale genetic screens.

But many leads have already emerged, according to Jill Heemskerk, the project organizer at NINDS. The investigators are currently keeping quiet about which compounds have shown good activity — but



Active approach: postdoc Federica Piccioni joins in the NIH's search for neurological agents.

some are already being prepared for follow-up studies in mice, says Heemskerk.

NINDS hopes that some of the hits will eventually move into clinical trials. If so, the trials could be completed rapidly as many of the compounds already have FDA approval, says Heemskerk. The concept might also be extended to other diseases, she adds. ■

Badgers set for cull in resumed tuberculosis trial

David Adam, London

Trials in which thousands of badgers will ultimately be killed to investigate their role in spreading bovine tuberculosis (TB) have



Thousands of badgers may die as a study into bovine tuberculosis is resumed in Britain.

resumed in Britain amid fresh arguments over the study's aims and effectiveness.

The trials — suspended early last year because of the country's outbreak of foot-and-mouth disease — are intended to show whether strategic culls of badgers could combat the spread of TB in cattle, which is a growing problem in parts of Britain.

"It is a tragedy that these trials are resuming when a growing body of scientific evidence indicates that cattle-to-cattle transmission is the major cause of the increase in bovine TB," says Elaine King, a zoologist and chief executive of the National Federation of Badger Groups.

Other scientists say that the situation is not so straightforward. "Badgers are part of the problem and continuing these trials is the only way to quantify how big a part they play," says Tim Roper, an animal-behaviour researcher at Sussex University.

The trials are designed to compare three strategies: killing all badgers in an area;

removing them only if a case of bovine TB is spotted; and no culling at all. But they have been dogged by delays and controversy since they were first proposed in 1997 (see *Nature* 394, 821; 1998). In a 1999 report, the House of Commons Select Committee on agriculture questioned whether the trials will produce statistically meaningful results.

Stephen Harris, a zoologist at Bristol University who runs the UK National Badger Survey and is a long-standing critic of the trials, describes them now as a "farce". Large numbers of cattle will have been replaced in trial areas since the outbreak of foot-and-mouth, he says, making it impossible to relate test results to the culling of badgers.

John Bourne, head of the government's Independent Scientific Group on bovine TB, which is running the trials, dismisses Harris's criticisms, saying that only about 5% of farms in trial areas were affected by foot-and-mouth, and that trial results — due in 2005 — will still be valid. ■

NINDS

V. LATTORD/NFEG

Search for martians 'too risky' if there's any sign of life

Washington Travelling 50 million miles to Mars is risky enough. Now a National Academy of Sciences report has detailed the environmental risks awaiting astronauts when they get there.

Among them are dust that sticks to electronic equipment, and toxic metals in the soil, including hexavalent chromium — a substance made famous by the film *Erin Brockovich*, in which the metal poisons a small town's drinking water. The report says that experiments to test this risk, and others to measure the radiation dose in tissue-equivalent material at the martian surface, should be carried out by precursor missions.

To guard against "back-contamination" of Earth, the study panel warns against sending humans to any site not previously shown to be free of organic material, even though "the committee recognizes that this may be in conflict with one of the primary goals of the exploration of Mars: to find extraterrestrial life".

The report also says that NASA should consider how to ensure a safe landing on the planet, and recommends that the

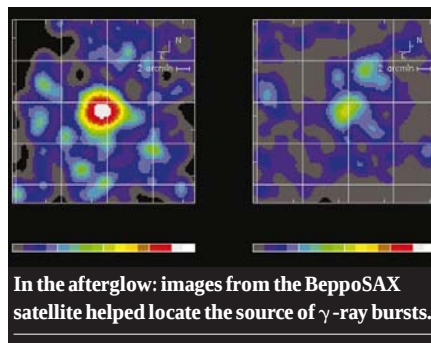
agency develops a high-resolution, three-dimensional map of any potential landing sites.

♦ www.nap.edu

Ageing satellite shut down amid fears of big bang

Rome The Italian–Dutch satellite BeppoSAX was shut down last week after engineers warned that it could be about to explode.

BeppoSAX, an X-ray-imaging spacecraft launched in 1996, had already delivered data for three times its planned lifespan. But time was slowly taking its toll on the satellite. When a fault was detected in the thermal control of the craft's battery, the Italian Space Agency decided to send a close-down command rather than risk an explosion,



which would have created a cloud of debris in the satellite's low-Earth orbit. BeppoSAX will eventually burn up in the Earth's atmosphere as its orbit decays.

The satellite is best known for its 1997 image of an X-ray afterglow of a γ -ray burst. The image helped to pinpoint the distant source from which the γ -rays originated. Such bursts are now thought to come from the fireballs that surround supernovae.

Prime candidate quits race to head Livermore

San Diego A leading candidate for the directorship of the Lawrence Livermore National Laboratory, the Californian nuclear-weapons research centre, has withdrawn from the troubled selection process.

Ray Juzaitis, a nuclear engineer and associate director at Los Alamos National Laboratory in New Mexico, said he was pulling out of the race after press reports made "an unwarranted link" between him and the Wen Ho Lee affair. Lee, a former employee under Juzaitis at Los Alamos, fell under suspicion of spying for China in 1999.

Juzaitis wrote on 30 April to the University of California, which manages Livermore for the US Department of Energy,

stating that the “unfounded controversy” could hinder the effectiveness of his proposed leadership.

Three other candidates are still believed to be in contention: Michael Anastasio and Jeffrey Wadsworth, both deputy directors at Livermore, and Steve Koonin, provost of the California Institute of Technology.

Britain urged to dump nuclear-waste agencies

London Britain’s nuclear-waste management agencies should be scrapped and replaced with independent and more transparent bodies, according to a new report from the Royal Society.

The report says that the public has lost confidence in bodies such as Nirex, a company set up by the government in the early 1980s to advise on waste management. Nirex has close links to the nuclear industry, and has been accused of suppressing research relating to new waste-storage sites.

Money should be spent on developing advanced materials to encapsulate high-level nuclear waste, such as plutonium and mixed-oxide fuel, before disposal, the report says. Britain’s nuclear-waste research base has been “seriously eroded”, says Geoffrey Boulton, a geochemist at the University of Edinburgh who chaired the panel that produced the report.

Decisions on building future nuclear power plants should remain on hold until the waste-storage problem is addressed, the society adds.

Edison’s fit of pique lit the way for *Science*

Washington Thomas Edison felt none too kindly about *Nature*, and so helped to set up *Science*, according to a biography of the US inventor published earlier this year.

Shortly after the 33-year-old Edison demonstrated an incandescent bulb in 1879, *Nature* published an article alleging that he was “thirty-five years behind the time in his new invention”, citing an 1845 patent for a similar device held by the English inventor Edward King.



Thomas Edison and one of his bright ideas.

After reading the comment, the enraged Edison decided secretly to underwrite *Science*. The journal published its first issue in 1880, and shortly thereafter wrote of Edison: “While others have talked he has worked ... those who are competent can see and judge for themselves”, according to R. V. Alvarado’s *Critical Lives: The Life and Work of Thomas Edison* (Alpha Books, 2002).

Sex shock as fetal error gives cloners bull fright

São Paulo Note to cloning researchers: always check the label on samples before beginning an experiment.

The dangers of getting samples mixed up were exposed by scientists at the University of São Paulo in Brazil. The birth of a cloned bull calf stunned researchers there, as they were expecting a female calf. They thought they had created a cloned embryo using cells from a cow’s ear. But genetic profiles last week showed that male fetal cells that were stored in the same freezer had been used instead.

“It was a surprise,” admits Jose Visintin, head of the cloning project. The team intended to clone both the adult and fetal cell lines to explore differences in chromosome ageing between the resulting clones. Two other expectant cows, due in August and January, are now expected to bear bulls too.

BETTMANN/CORBIS

The state of the planet

More than 2,000 experts will be involved in a four-year effort to survey the health of the world's ecosystems and threats posed by human activities. Virginia Gewin profiles the Millennium Ecosystem Assessment.

In March last year, Harold Mooney scribbled a flow chart on a napkin as he and Walter Reid sat in a Montreal bar. The magnitude of the task they had embarked on was really starting to sink in.

Instead of evaluating how ecosystems respond to just one environmental concern, such as climate change, they were talking about providing a complete planetary health check, determining the impacts of changes in land use, loss of biodiversity, the application of agricultural fertilizers, and many other factors — a truly colossal endeavour.

Reid, formerly a zoologist with the World Resources Institute (WRI) in Washington DC, and Mooney, an ecologist at Stanford University in California, are now executive director and assessment-panel co-chair, respectively, of the Millennium Ecosystem Assessment (MEA). This is a US\$21-million, four-year effort to determine the state of the Earth's ecosystems that will seek input from more than 2,000 leading natural and social scientists. Its financial backers include

the David and Lucile Packard Foundation, the United Nations' Global Environment Facility, media mogul Ted Turner's UN Foundation, and the World Bank.

Reid first conceived the project as a proposal for one of the biennial reports on the environment compiled by the WRI, together with the UN Development Programme,

the UN Environment Programme (UNEP) and the World Bank. The result was the Pilot Analysis of Global Ecosystems (PAGE), which provided the technical underpinning for the *World Resources 2000–01* report, released in June last year. PAGE, which involved 500 contributors, concluded that the capacity of ecosystems to meet human needs for food and clean water is diminishing, and warned of threats to biodiversity and human health.

Need to know

But Reid's brainchild had by then outgrown the scope of a WRI report, and in July 2001 the MEA was launched. Ecologists and policy-makers had been murmuring about the need for a comprehensive planetary health survey for some time. With 60% of the world's major fisheries being overfished, some 14 million hectares of forest disappearing each year, and habitats from wetlands to coral reefs under threat, credible guidance on how to manage these resources is invaluable.

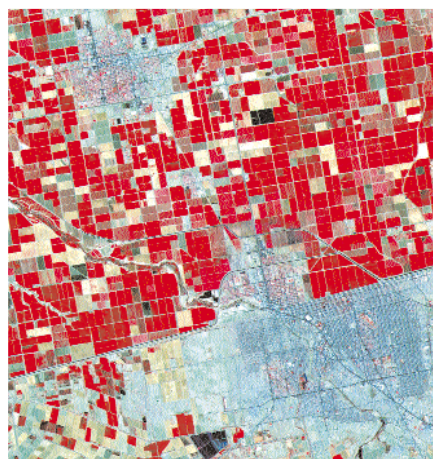
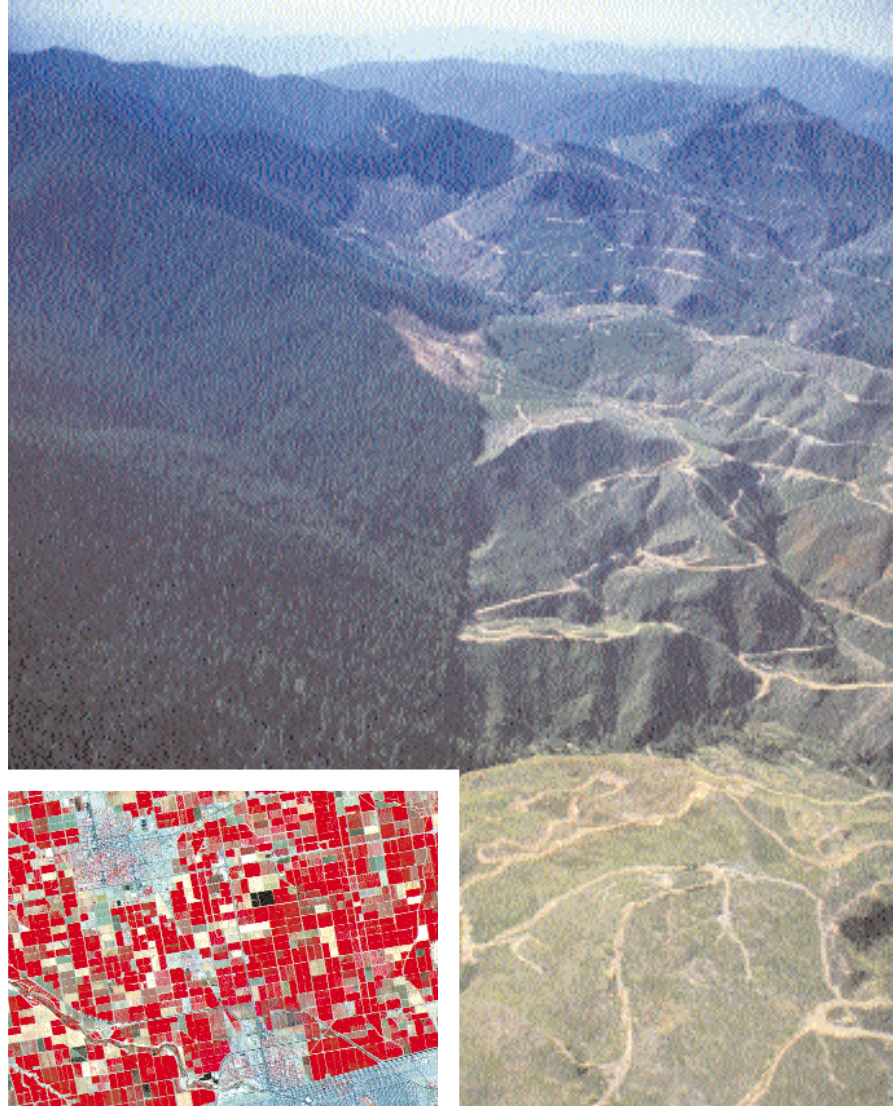
Ecosystems are often managed to obtain one product or service, often at the expense of others. For example, increasing the food sup-

ply can entail converting forest to agriculture, which can reduce biodiversity and the supply of timber and clean water. But no one has previously tried to work out how all of these conflicting pressures interact. "The trade-offs and interactions are crucial," says Mooney.

Crucial, but also fiendishly complicated. Arguments about how to deal with scientific uncertainties have dogged the Intergovernmental Panel on Climate Change (IPCC), often cited as a model for the MEA. But in some respects, the climate panel had it easy. Baseline data on the oceans and atmosphere are reasonably good, as are the global circulation models used to examine how the climate system will respond to different atmospheric levels of greenhouse gases.

The MEA enjoys no such luxury. In many cases, data on ecosystems are sketchy, and models of how human activities affect hydrology, biogeochemistry and biodiversity are not so well developed. Even tougher is trying to determine how changes in ecosystems will influence human well-being. "This is about five times as big as the IPCC," concludes Prabhu Pingali, an economist at CIMMYT, the International Maize and Wheat Improvement Center in Texcoco, Mexico, and co-chair of the MEA's Scenarios Working Group.

Groundwork: satellite images will reveal deforestation and other changes in land use.



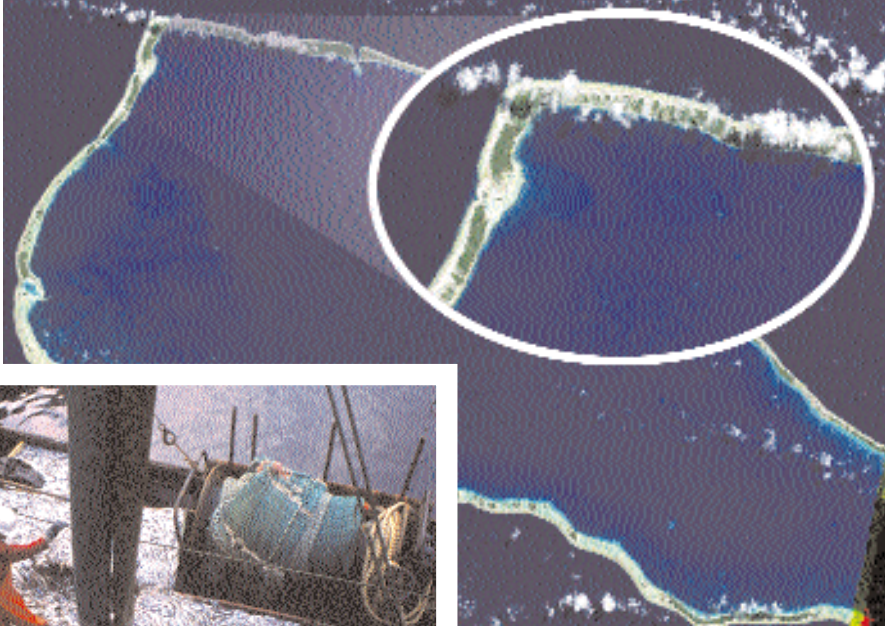
D. DANCER/STILL PICTURES

NASA/GSFC/MITI/ERSDAC/JAROS/ASTER

C. REID



Walter Reid is now the assessment's director.



All at sea: overfishing and coral bleaching threaten the health of marine ecosystems.



M. EDWARDS/STILL PICTURES

Having established its secretariat at ICLARM, the World Fish Center in Penang, Malaysia, in January this year, the MEA is now gathering existing data on ecosystems and the services they provide — such as food, water purification, nutrient cycling and income from tourism. Volunteer scientists are creating a baseline data set that will detail the condition and geographical extent of the world's terrestrial, marine and freshwater ecosystems.

They will have to trawl far and wide. Remote-sensing data will come from donated or public sources. The US government, for instance, has made land-use data from a variety of satellite projects available to the MEA. Smaller research projects, national governments and organizations such as the UN Food and Agriculture Organization (FAO) and the World Conservation Union also possess vital data on demography, biodiversity and agriculture. Any relevant peer-reviewed science will be thrown into the mix. Data on road networks — a key factor in environmental degradation — will also be crucial.

Like the IPCC, the MEA will seek to forecast potential outcomes for a variety of scenarios. Pingali's group is currently working on as many as four story-lines. Narrowing down the range of

predictions — optimistic, pessimistic and variations in between — is no small task. Pingali and his colleagues must first identify plausible scenarios for demographic, technological, economic and social changes. They will then use this information to forecast changes in pollution, land use and other key variables.

The parallels with the IPCC are legion: not only has the MEA followed the climate panel's division into specialist working groups, it has also snared Robert Watson, who until last month was chair of the IPCC, to serve as co-chair of the 40-member board that oversees the MEA's work. Watson, the World Bank's chief scientist, is widely credited with facilitating the scientific consensus on climate change that emerged in the IPCC's reports. "It is important to show a range of views, not just one view," he says. "You must explicitly say what is known and what is not known."

Avoiding meltdown

Defining where science ends and policy begins will be difficult. The MEA aims to detail the likely implications of different policy options, but its leaders want to avoid any perception that they are urging the adoption of specific policies. This is wise, agrees William Clark, a Harvard University ecologist who leads the Global Ecosystem Assessment, an effort to determine how the MEA and smaller assessments can be most effective. "It's when science tells policy-makers what to do that you have a meltdown," he says.

The IPCC's success shows that it is possible to get the balance right. But the MEA differs from the climate panel in one important respect: whereas governments asked scientists to form the IPCC, the MEA is a grassroots initiative from scientists who saw a need to fill the world's ecological data gaps.

Reid is aware of the difficulty of getting governments to pay attention to reports that they haven't commissioned. In March 1992, he attended a preparatory meeting in New York for the UN Conference on Environment and Development, held in Rio de Janeiro later that year. At that meeting, Wen Lian Ting, Malaysia's representative to the FAO and UNEP, rallied developing nations against the Global Biodiversity Assessment (GBA), a biodiversity inventory proposed by scientists and UNEP to provide support to negotiations

for the UN Convention on Biological Diversity (CBD). As countries preparing to sign the convention, which was launched in Rio, had not specifically requested the GBA, Ting viewed it as another example of the rich North telling the developing South how to use its resources.

The scientists involved in the GBA completed their work anyway — but the hostility stirred up by Ting ensured that its report has remained a reference work, rather than a genuine contribution to the political process. "I don't think any of the scientists really understood how significant it was that governments didn't accept the GBA," says Reid.

Having learned this lesson, the MEA's leaders have worked hard to win political legitimacy, seeking authorization from relevant international conventions. Signatories to the CBD, the Convention to Combat Desertification and the Ramsar Convention on Wetlands have each endorsed the MEA. Following a request from the CBD, the MEA will include specific chapters on biodiversity.

Questions and suggestions

Reid says that Ting kept popping up in his thoughts, causing him to ask: "What would it take for her to accept it?" With this in mind, MEA officials have courted individual governments, listening to their suggestions. They have also made the project more attractive to governments by conducting local assessments alongside the global effort. Six of these 'subglobal' assessments — for western China, Norway, Sweden, India, southern Africa and Papua New Guinea — are already scheduled. Most are bankrolled by the countries involved, but the MEA offers seed funds for surveys in developing nations, and the Papua study is funded by the UN Development Programme. More subglobal assessments may be added, as long as they follow technical guidelines and engage target users.

Industry has also been involved in the MEA's consultations, and its board includes Anthony Burgmans, chief executive of the Anglo-Dutch consumer-products manufacturer Unilever. "For this to work it has to have a broad acceptance in society," he says.

Whether the MEA will work will not be known for several years. Its success will be judged on whether it can gain the high credibility enjoyed by the IPCC and become the *de facto* scientific advisory body for the conventions that have endorsed its work.

For the time being, the reactions of the key players to the MEA's progress are positive. "The MEA is a timely and worthwhile exercise which needs the support of all who demand a fair, honest and constructive appraisal of the state of our natural resources," says Ting. Coming from a self-proclaimed sceptic who has shown that she can make or break such efforts, that is a significant endorsement. ■

Virginia Gewin is an intern with *Nature*.
www.millenniumassessment.org

STANFORD NEWS SERVICE



The task facing Harold Mooney is immense.

Magnetic mind games

Using short magnetic pulses, neuroscientists are reaching into the human skull and temporarily altering volunteers' brain activity. Marina Chicurel takes an induction course.

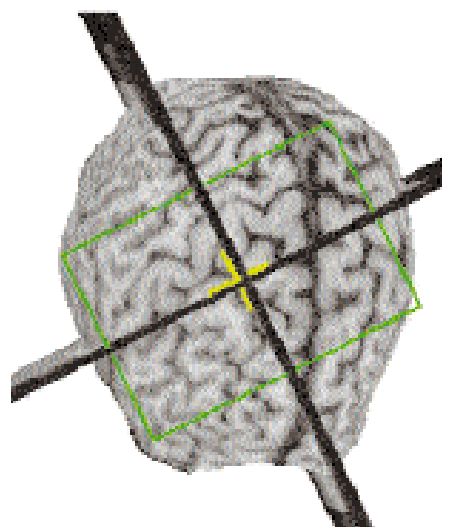
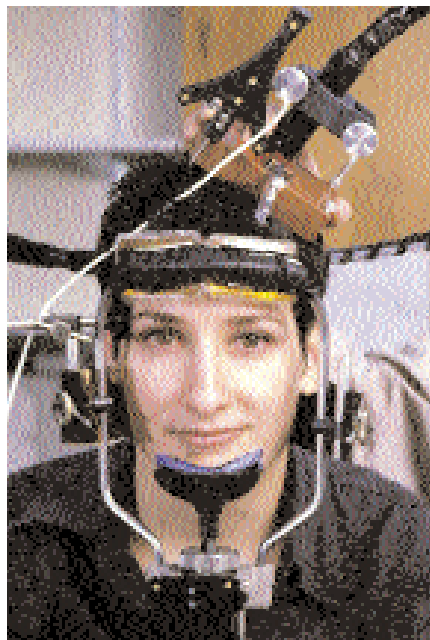
Studying the human brain can be a frustrating business. Although sophisticated imaging techniques can offer snapshots of activity, direct intervention in the brains of humans is ethically off-limits. It is no wonder that neuroscientists sometimes feel like visitors to a museum — they can look but not touch.

But things are beginning to change. Neuroscientists studying human cognitive processes are now making tentative use of a technique that can temporarily alter their subjects' brain activity. Insights into everything from language to conscious awareness are emerging, as well as hints that the technique could help treat some mental disorders. "There are just some extraordinary things coming out," says Michael Gazzaniga of Dartmouth College in Hanover, New Hampshire, editor-in-chief of the *Journal of Cognitive Neuroscience*.

Before technology for imaging the human brain emerged, neuroscientists interested in cognitive functions were mostly limited to studying patients with brain lesions — damage to specific areas of the brain. By looking for links between the patients' symptoms and the site of the damage, researchers pinpointed brain areas involved in abilities such as language,

learning and memory.

During the 1980s and 1990s, techniques such as functional magnetic resonance imaging (fMRI), which tracks brain activity by monitoring blood flow, allowed neuroscientists to see which areas of the brain were active during specific tasks. A flood of information about the functions of



A century-old dream to influence brain activity with magnets (below), is now a reality using powerful coils (left) and imaging (above).

different brain areas followed.

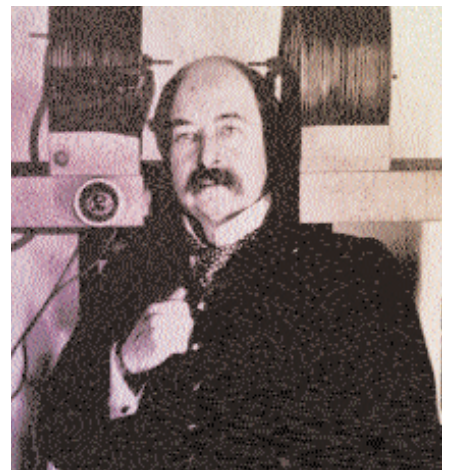
But such techniques have their limits. "There's no causality in it," complains Tomás Paus, a neuroscientist at McGill University in Montreal, Canada. "You can never say whether one brain region is influencing another, or whether they are active together because some third region is driving both." Rather than just passively observing the brain, researchers would like to manipulate it directly.

Polar explorers

Scientists had long suspected that magnetic fields might allow them to do this. Brain cells send electrical signals along the fibres that make up their communication networks. Because changing magnetic fields can induce current in electrical conductors, researchers thought that a magnetic pulse might stimulate currents in brain cells, and so alter brain activity. Early attempts to do this date back to the nineteenth century, but it wasn't until 1985 that devices capable of producing the short, intense pulses needed to stimulate the brain were developed.

The breakthrough was made by Anthony Barker, a medical physicist at the Royal Hallamshire Hospital in Sheffield, UK. Barker used a 2-tesla magnetic pulse about 1 millisecond long to stimulate the brain area that controls finger movement. Unintentionally, the volunteer's fingers twitched¹. "It caused a good deal of excitement," recalls Barker.

The technique, known as transcranial magnetic stimulation (TMS), remains similar today. There are hints that the intensity and frequency of the induced current can influence whether the stimulation increases or damps down brain activity^{2,3}. But TMS cannot, at present, exert precise control over the brain. It puts in "random neural noise", explains Vincent Walsh of the University of



Oxford, UK, who has used TMS to study cognitive functions such as visual awareness.

Single pulses affect brain function for just a few milliseconds. But a train of pulses, typically delivered at rates of about one per second, can disrupt brain activity for tens of minutes. This technique, known as repetitive TMS (rTMS), creates 'virtual' lesions for neuroscientists to experiment on.

Last year, for example, Alfonso Caramazza and colleagues at Harvard University used rTMS to reveal how our brains seem to use different regions to handle verbs and nouns. Caramazza was studying the left prefrontal cortex, an area thought to be involved in the ability to conjugate verbs. Previous lesion studies had proved inconclusive, partly because the damage often extended beyond the area of interest. And, as with other studies of brain damage, it was impossible to gauge the extent to which other parts of the patients' brains had compensated for the damage.

"With brain-damaged patients we're at the mercy of nature," says Caramazza. "That's why I became a TMS convert." When Caramazza applied rTMS to the left prefrontal cortex he found that his subjects had difficulty conju-



Anthony Barker used magnets to control movement.



Moving story: Alvaro Pascual-Leone believes TMS holds the key to unravelling many mental processes.

gating verbs⁴. But their ability to give singular and plural forms of nouns was unchanged.

In the past few years, rTMS has been used to investigate a number of thought processes. In 1999, for instance, Stephen Kosslyn, together with Alvaro Pascual-Leone of the Beth Israel Deaconess Medical Center in Boston, led a team that examined the long-standing hypothesis that when we visualize a scene in our mind's eye, we generate mental 'images' that recreate the relative distances between objects in the real scene. Support for the idea came from studies of a brain region known as V1. When we view a real scene, the pattern of active neurons in V1 forms a kind of 'map' of the scene, and some studies indicated that the same area was active during visualizations⁵.

But other experiments had failed to confirm this. It was also unclear whether the activation was intimately involved in visualization, or was simply a by-product of activity in other brain regions. Kosslyn asked volunteers to memorize a pattern of stripes, and then close their eyes and answer questions about the pattern. Applying TMS to V1 increased

the time the subjects took to answer⁶, supporting the idea that our mind's eye conjures up lifelike images.

Other researchers have recently used TMS to study visual attention⁷, the storage and retrieval of memories⁸, and how we recognize our own face⁹ or the angry expressions of others¹⁰. As interest in the technique grows, researchers are exploring ways of using it in conjunction with imaging technology.

Image conscious

Some groups are combining TMS with fMRI. A reflective tag, which can be tracked by an optical sensor, is attached to the coil, allowing the coil's position to be displayed on the MRI scan. This and similar methods should allow the neural connections within the brain to be mapped out. Paus, for example, is interested in probing changes associated with disease, and plans to examine connectivity in the frontal cortex of schizophrenic patients, which some researchers believe may be abnormal. Other experiments, which are in their early stages, have

been designed to investigate how tasks such as learning change neural connections.

TMS can also be used to probe the function and timing of the signals that travel along neural pathways. Last year, Walsh and Pascual-Leone examined connections between V1 and another area in the visual cortex known as V5. Visual information from our eyes arrives at V1 and is relayed to V5, which plays a specialized role in the perception of motion. V5 also sends signals back to V1, and some researchers believe that this connection makes us aware of the motion detected in V5.

Earlier work had shown that magnetic stimulation of V5 caused subjects to see moving spots of light. So Pascual-Leone and Walsh stimulated V5 and then used a less intense pulse to disrupt activity in V1. Stimulating both areas simultaneously had no effect on what subjects reported seeing. But the appearance of the moving spots was greatly impaired if the V1 pulse was delivered between 5 and 45 milliseconds after the V5 stimulation¹¹. The results show that the link from V5 back to V1 plays a role in motion awareness, as well as demonstrating that this link operates very quickly.

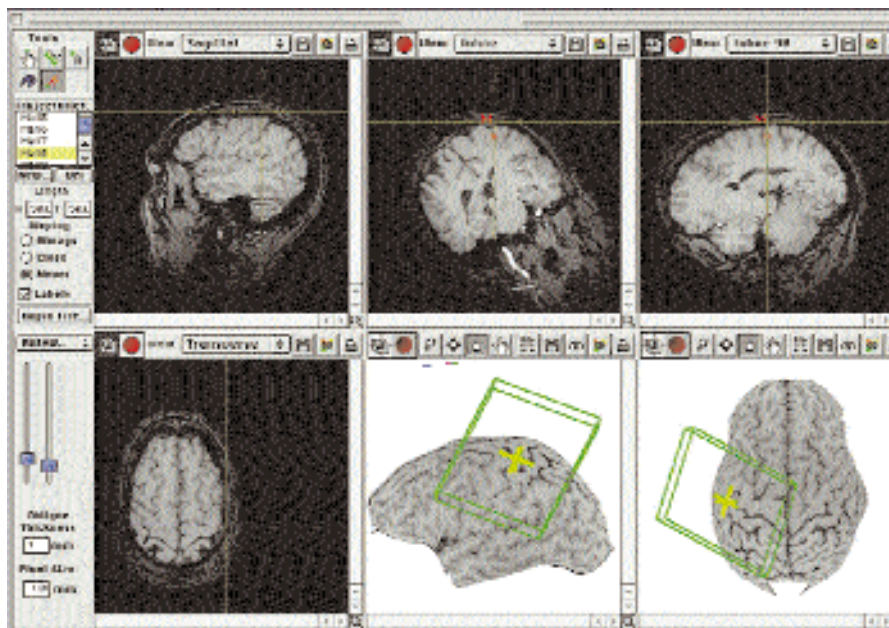
Slow headway

Walsh says that many more timing issues could be probed using similar studies. "It's amazing how many easy experiments haven't yet been exploited," he says. But TMS is taking a long time to be adopted by cognitive scientists. One factor may be the relatively poor communication between cognitive neuroscientists and those studying movement — the first group to embrace the technique. Some researchers also worry about the spatial resolution of TMS. The technique is thought to stimulate about one cubic centimetre of brain, but factors such as coil position and pulse intensity influence the affected volume in a way that is not clearly understood.

But some may have been deterred by fears that TMS could be harmful. "A concern is how virtual is the virtual lesion, especially in subjects who may have some unidentified vulnerabilities, such as an undiagnosed tendency to have seizures," says Douglas Rothman, a medical-imaging expert at Yale University in New Haven, Connecticut.

But TMS does have a good safety record. Current guidelines for the length and intensity of the stimuli were drafted in 1998 following a consensus reached at an international workshop two years earlier¹². Some subjects have suffered single episodes of rTMS-induced seizures. But, according to Pascual-Leone, only seven such cases have been reported, none of which occurred when the published guidelines were followed.

Pascual-Leone also notes that most researchers give their subjects cognitive tests to make sure that they are back to normal before they leave the lab. But Jordan Grafman



Precise picture: imaging techniques allow the magnetic stimulus to be applied in a specific position.

of the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland, thinks that these evaluations should be standardized to ensure that they pick up subtle effects. "You might want to have people stay longer, you might want to have additional testing," he says. "There have been very few studies considering this."

Despite these questions, most agree that isolated sessions of rTMS, such as those experienced by volunteers in cognitive studies, do not induce permanent changes. But repeated bouts of rTMS can have long-lived effects, and some hope to harness them to treat mental disorders. In 1995, for example, after noticing that depressed patients had low levels of activity in their left prefrontal cortices, Mark George, a psychiatrist at the Medical University of South Carolina in Charleston, tried boosting activity in that region using rTMS¹³. George says that his patients improved, but similar studies conducted since have proved to be less clear-cut.

A team led by Tal Burt, a psychiatrist at Columbia University in New York, has

recently collated the results of 25 such studies¹⁴. The group concludes that rTMS has a modest antidepressant effect, of uncertain clinical value, but which could improve as researchers refine their methods. Factors such as coil placement, stimulation frequency and the influence of patient variability now need to be investigated. George is planning large-scale trials. "The questions now are: how big is that effect, can it be made to last, is it clinically useful, what are the use parameters that maximize the effect?" he says.

Virtual surgery

Other disorders are also being tackled with rTMS. Last year, Massimiliano Oliveri, a neuroscientist at the University of Palermo in Italy, working with colleagues at the University of Turin, showed how the technique can be used to treat hemispatial neglect, a condition in which patients have difficulty paying attention to objects on one side of their visual field.

The condition is often caused by damage to parts of the brain called the parietal cortex. Under normal circumstances, the left and right sides of this area inhibit each other. But when one side is damaged, the other becomes hyperactive. Although, to a degree, the brain may compensate for this, Oliveri reasoned that he might be able to balance the activity by creating a virtual lesion in the undamaged side. Initial results were promising, with the patients showing improved awareness during stimulation¹⁵, and Oliveri is now using stimuli designed to produce longer-lived changes — 1 Hz for 10 minutes every day for five days. The eight patients he has treated, and followed for a month so far, have improved 40% faster than non-treated controls, he says.

TMS may also be able to go one step further, and enhance normal peoples' thought

processes. Last year, Grafman applied rTMS to normal volunteers while they solved a reasoning puzzle¹⁶. "We decided to see if we could actually make people faster," he says. By applying rTMS to the prefrontal cortex, an area that is active when people solve certain types of visual puzzle, Grafman reduced the time taken by subjects to solve a problem involving geometric shapes.

But it is unclear why such experiments work. Walsh says that noise can boost performance when it is used to block an inhibiting area, such as in Oliveri's studies of neglect, but it is difficult to see how it could enhance a circuit's precisely tuned pattern of activity. Grafman acknowledges this apparent paradox, but points out that noise can, in some systems, aid a change from one state to another — a phenomenon known as stochastic resonance.

Some scientists are uneasy about this area of research. Pascual-Leone believes it might be possible to use TMS to enhance a wealth of mental skills. "But those experiments shouldn't be done without a very serious debate," he says. "And who is to decide what behaviour should be modified, and by whom, and for what purpose?"

Others do not see an ethical dilemma in such experiments. George has been funded by the US Department of Defense to study whether TMS can help to improve memory. He argues that such methods are no different from accepted means of boosting performance. "We do things all the time to try to enhance our performance," says George. "We exercise, we practice. The idea of using a physical stimulation to do that doesn't seem to me to broach any new ground."

Whether or not such controversial projects succeed, TMS enthusiasts see a bright future for the technique. After more than a decade of relative obscurity, they say TMS is beginning to take its place in the toolboxes of cognitive neuroscientists. "People have started realizing how many unique opportunities magnetic stimulation offers," says Pascual-Leone. "I think they've started to get it."

Marina Chicurel is a writer in Santa Cruz, California.



Vincent Walsh believes that there are many easy experiments involving TMS waiting to be done.

1. Barker, A. T., Jalinous, R. & Freeston, I. L. *Lancet* **1**, 1106–1107 (1985).
2. Chen, R. et al. *Neurology* **48**, 1398–1403 (1997).
3. Maeda, F., Keenan, J. P., Tormos, J. M., Topka, H. & Pascual-Leone, A. *Clin. Neurophysiol.* **111**, 800–805 (2000).
4. Shapiro, K. A., Pascual-Leone, A., Mottaghy, F. M., Gangitano, M. & Caramazza, A. *J. Cogn. Neurosci.* **13**, 713–720 (2001).
5. Mellet, E., Petit, L., Mazoyer, B., Denis, M. & Tzourio, N. *Neuroimage* **8**, 129–139 (1998).
6. Kosslyn, S. M. et al. *Science* **284**, 167–170 (1999).
7. Fierro, B. et al. *NeuroReport* **11**, 1519–1521 (2000).
8. Rossi, S. et al. *Nature Neurosci.* **4**, 948–952 (2001).
9. Keenan, J. P., Nelson, A., O'Connor, M. & Pascual-Leone, A. *Nature* **409**, 305 (2001).
10. Harmer, C. J., Thilo, K. V., Rothwell, J. C. & Goodwin, G. M. *Nature Neurosci.* **4**, 17–18 (2001).
11. Pascual-Leone, A. & Walsh, V. *Science* **292**, 510–512 (2001).
12. Wassermann, E. M. *Electroencephalogr. Clin. Neurophysiol.* **108**, 1–16 (1998).
13. George, M. S. et al. *NeuroReport* **6**, 1853–1856 (1995).
14. Burt, T., Lisanby, S. H. & Sackeim, H. A. *Int. J. Neuropsychopharmacol.* (in press).
15. Oliveri, M. et al. *Neurology* **57**, 1338–1340 (2001).
16. Boroojerdi, B. et al. *Neurology* **56**, 526–528 (2001).

Basic or applied, it's the interaction that counts

Toxicology must not be driven by politics: it's science that should drive decision-making.

Sir—While we sympathize with some of the views expressed by Marcello Lotti and Pierluigi Nicotera in their Concepts essay “A risky business” (*Nature* **416**, 481; 2002), the authors fail to acknowledge the wider role of toxicology in society. Toxicology is much more than the purely mechanistic studies they propose, and should not be incorporated into “the mainstream of fundamental biomedical research”. Rather, it should draw from and interact with such research.

Toxicology has several functions: to identify the origin and nature of toxic insults to human health; to identify their mechanism of action; and to guide regulatory and public-health bodies in measures to remove or reduce toxicological threats. Although mechanistic studies are essential, it is just as important to know the nature of the compounds or environmental insults responsible for adverse effects, and the dosimetry of such exposures. This enables effective risk assessments and development of strategies to reduce risks to human health. Lotti and Nicotera fail to cite any example where mechanistic knowledge alone has allowed reliable risk predictions for humans.

Contrary to the authors' claim that basic research has become irrelevant to many toxicologists (partly justified in the past), recent national and international meetings demonstrate that toxicology has embraced and contributed to the advances of basic molecular and cell biology, as well as incorporating pharmacology, pathology, chemistry and epidemiology. Toxicology is a multidisciplinary subject, and there will always be situations which demand urgent action, irrespective of our understanding of basic biological mechanisms.

If the strategy proposed by Lotti and Nicotera had been applied to liver cancer in Southeast Asia and Africa, it would certainly have produced some interesting mechanistic data on hepatocarcinogenesis. But it was the integration of molecular dosimetry and epidemiological studies that identified aflatoxin B1 as a major risk factor for this cancer. Subsequently, the knowledge of carcinogenic mechanisms provided the rationale for measures to reduce the risk, such as improved food storage, and the design of chemoprevention trials. Similar remarks could be made about organophosphorus pesticides and neurotoxicity; polycyclic aromatic hydrocarbons, aromatic amines and carcinogenicity; accidental or deliberate chemical poisoning; the development of antidotes; and the relevance of endocrine

disruptors to human health. In all these it is essential to carry out ‘chemical-driven’ as well as ‘mechanism-driven’ toxicology, using cutting-edge methodologies, to generate meaningful risk assessments.

The rapidly expanding knowledge about genetic polymorphisms of enzymes involved in activation (such as cytochrome P450) and detoxification (for example glutathione S-transferase and *N*-acetyl transferase) of toxins shows that some population groups may be particularly susceptible to some chemicals. The mechanistic study of polymorphisms is valuable, but without considering the nature of the compounds to which individuals are exposed, such knowledge is of little practical value. The public is more concerned to know what compounds are toxic and to see them removed from the environment than to know in detail which biochemical mechanisms are occurring in their cells, important though that may be.

Society rightly demands assurances on

the safety of the environment, and this can be achieved only by the application of toxicological principles together with mechanistic understanding. Nevertheless, we should strive to ensure that political agendas do not drive toxicology, but rather that strong science should drive regulatory decisions. It is the interaction between fundamental and applied science that is important in toxicology, not one or the other component.

Peter B. Farmer

Biocentre, University of Leicester, University Road, Leicester LE1 7RH, UK

Other signatories to this letter:

Helmut Bartsch German Cancer Research Centre, Heidelberg
Alan Boobis Imperial College, London, UK

Kevin Chipman School of Biosciences, University of Birmingham, UK

Andy Gescher Department of Oncology, University of Leicester, UK

Fred F. Kadlubar National Center for Toxicological Research, Jefferson, Arkansas, USA

Margaret M. Manson Biocentre, University of Leicester, UK

David H. Phillips Institute of Cancer Research, London, UK

David E. G. Shuker Department of Chemistry, Open University, UK

James A. Swenberg University of North Carolina, USA

Steven R. Tannenbaum Massachusetts Institute of Technology, USA

Christopher P. Wild University of Leeds, UK

Fundamentals are still relevant in toxicology

Sir—Marcello Lotti and Pierluigi Nicotera, in their Concepts essay on toxicology (*Nature* **416**, 481; 2002), raise important issues about the seminal contribution made by the use of toxins as probes to our understanding of fundamental principles and functions of biological systems. But their claim that “basic research has, over the past two decades, become irrelevant to many toxicologists” is not true.

As the current and incoming presidents of the Society of Toxicology, a professional organization of more than 5,000 members, we can attest that our society has worked diligently to undo that outdated image of toxicology, and has taken up the challenge that “Toxicology research should urgently appraise its performance and join mainstream biomedical science”. This is reflected in the quality of science presented at the society's meetings and published in leading journals. As part of a long-range plan to guide research, education and outreach (see www.toxicology.org), our priorities are to facilitate research with an emphasis on mechanisms of action, and to increase the use of relevant science in risk assessment and decision-making.

Lotti and Nicotera state “Toxicology is being shaped by worldwide political agendas, triggered by the public's desire for swift and precautionary solutions to the

possible health effects of environmental chemicals.” Risk assessment of chemical hazards is fraught with uncertainty, and the science of toxicology is sometimes misrepresented. It is in the public perception of hazards and health risks, and the blurred distinction between science and policy (see Roger A. Pielke's Commentary: *Nature* **416**, 367–368; 2002) that toxicology receives a ‘bad rap’.

Many factors go into risk-management decisions, and toxicology is but one piece of a complicated and highly political process managed by others with limited understanding of fundamental science. Basic research alone will not resolve uncertainty or conflict in risk assessment.

Fostering basic research is a primary goal of our discipline. But building better ways to put fundamental knowledge into risk assessment and regulatory practice is also vital. We believe that using knowledge to benefit human health is a responsibility shared by the entire biomedical research community. The National Institutes of Health and other sponsors of biomedical research continue to emphasize that the results obtained at the laboratory bench need to translate to improvements in the health of the public at large.

David L. Eaton*, **William F. Greenlee†**

**Center for Ecogenetics & Environmental Health and School of Public Health, University of*

Washington, Seattle, Washington 98105-6099, USA

†CIIT Centers for Health Research, PO Box 12137, Research Triangle Park, North Carolina 27709, USA

Creating a bioinformatics nation

A web-services model will allow biological data to be fully exploited.

Lincoln Stein

During the Middle Ages and early Renaissance, Italy was fragmented into dozens of rival city-states controlled by such legendary families as the Estes, Viscontis and Medicis. Though picturesque, this political fragmentation was ultimately damaging to science and commerce because of the lack of standardization in everything from weights and measures to the tax code to the currency to the very dialects people spoke. A fragmented and technologically weak society was vulnerable to conquest, and from the seventeenth to the nineteenth centuries Italy was dominated by invading powers.

The old city-states of Italy are an apt metaphor for bioinformatics today. The field is dominated by rival groups, each promoting its web sites, services and data formats. Unarguably, this environment of creative chaos has greatly enriched the field. But it has also created a significant hindrance to researchers wishing to exploit the wealth of genome data to its fullest.

Despite its shaky beginning, the nation of Italy was eventually forged through a combination of violent and diplomatic efforts. It is now a strong and stable component of a larger economic unit, the European Union, with which it shares a common currency, a common set of weights and measures, and a common set of rules for national and international commerce. My hope is that bioinformatics will one day achieve the same degree of strength and stability by adopting a universal code of conduct along the lines I propose here.

Screen scraping: mediaeval torture

The promise and peril of the bioinformatics landscape is clear to any bench biologist attempting to mine the human genome for information on, say, a favourite genetic region. The online sources of these data each provide remarkable user interfaces and deeply interconnected data sets of great richness. Yet each interface is different, both in the subset of data presented and in organization. The researcher may find herself devoting as much time adjusting to differences in presentation of the data as she does actually thinking about them. The situation is worse when comparing a human gene to its orthologue in another species. This brings the model organism databases into play, each of which has its own type of user interface and format. (See us.expasy.org/alinks.html; www.stat.wisc.edu/biosci/genome.html; and mbcf.dfci.harvard.edu/cmsmbr/biotools/biotools10.html for an idea of the scale of the problem.)

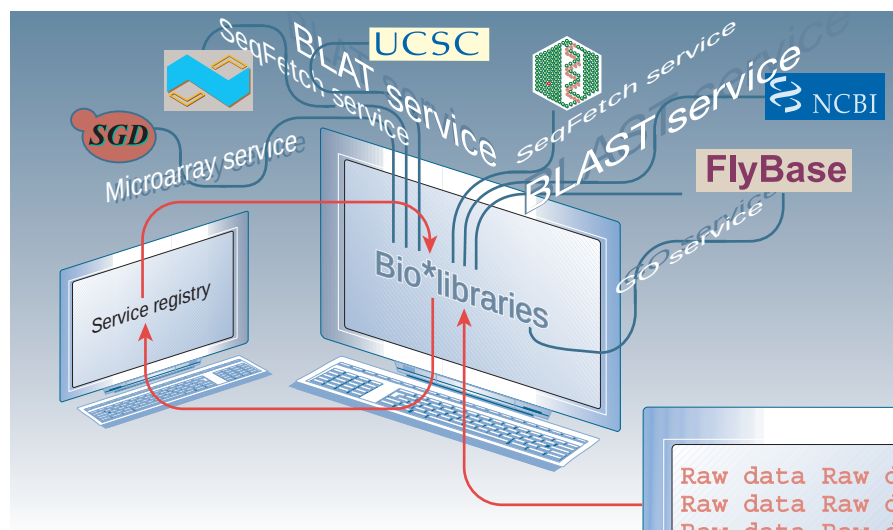


Figure 1 Moving towards a bioinformatics nation. Because each data provider (such as Flybase and UCSC) publishes data in an idiosyncratic form, the Bio* software package (Bio* libraries) was created to massage data into a standard internal format. Unfortunately, Bio* needs to be fixed each time a provider changes its formats. A web-services world would build on the successes of the Bio* projects by defining standard interfaces to various types of computations and data formats. The Bio* libraries can be written to recognize these interfaces, allowing them to interoperate easily with all data providers. A service registry would let data providers enter an electronic 'address book', allowing the Bio* libraries to locate and interact with new data sources automatically.

This inconvenience for the bench biologist is disastrous for the bioinformaticist, who typically needs to aggregate data from many online sources to create a data set for further analysis. When these data reside on different servers using different data formats and access methods, the first step is to write a set of software 'scripts' to fetch them, reformat them and place the extract into a local database. This is not straightforward, because most online biological databases were designed to be accessed by humans, not by machines. Bioinformaticists often find themselves writing scripts to parse the HTML source to extract the data while ignoring graphics links and explanatory text, a process called screen scraping.

Screen scraping is despised for various reasons. First and foremost, it is brittle. Database managers are always tinkering with the user interface, adding a graphic here, moving a button there, to improve the user experience. Each small change in a popular web page breaks dozens of screen-scraping scripts, causing anguished cries and hair-tearing among the bioinformaticists who depended on those scripts for their research and the research of the wet labs they support. Second, it is unreliable. There is no published documentation of what a data source's web pages are supposed to contain, so bioinformaticists must guess from a few

examples. Finally, there is massive duplication of effort. Almost every bioinformaticist has written a parser for the National Center for Biotechnology Information (NCBI) BLAST service at least once, sometimes many times. Because they are one-offs, these scripts are generally undocumented and not widely distributed. Most of them only work for a short time because BLAST changes every few months.

Bio* projects reduce the pain

The bioinformatics community has responded to the challenges posed by this 'city-state' situation with the Bio* projects, a series of freely available open-source projects (www.open-bio.org), in which nearly a hundred software engineers have developed re-usable code libraries in the Perl, Java, Python and Ruby programming languages (known as Bioperl, BioJava, Biopython and Bioruby, respectively). These libraries automate common bioinformatics tasks, such as manipulating DNA and protein sequences, and provide methods for importing and exporting data between data sources and among file formats. To fetch a piece of data from a database, the bioinformaticist uses the Bio* libraries to do the fetch, put the information in a standard format, and return the reformatted data to her script. This prevents duplication of effort. No one will ever again

be forced to write a BLAST parser.

But the Bio* libraries can't solve the problem of the brittleness of online data sources. As soon as one of these web pages changes, the Bio* library breaks and has to be patched up as quickly as possible. This works only because the Bio* libraries contain a series of adapter modules for each of the online databases, lovingly maintained by a group of dedicated (and very busy) programmers. The Bio* libraries also cannot immediately solve the problems of the data providers themselves. Whenever two providers need to exchange data, for example to share sequence annotation data, they must agree on an 'exchange of hostages' treaty in which they negotiate the terms and format of the exchange. Needless to say, this type of negotiation is awkward and time-consuming.

Strength through unity

To achieve seamless interoperability among online databases, data providers must change their ways. If they all had identical ways of representing biological data, standard user interfaces and standard methods for scripts to access the information, the problem of gathering and integrating information from diverse data sources would largely evaporate. But such conformity would destroy the creative aspect of online databases — and, indeed, no data provider would willingly surrender to it (but see Box 1 for a proposed code of conduct).

A more acceptable solution relies on the emerging technology of web services. A web service is a published interface to a type of data or computation. It uses commonly accepted data formats and access methods, and a directory service that allows scripts to find them. The web-services system shown in Fig. 1 allows several data sources to provide the same type of service so, for example, the NCBI, the University of California Santa Cruz (UCSC) and the European Bioinformatics Institute (EBI) could all provide a sequence-similarity service. A script written to use one service would work with them all, even though the three sites could implement the service in radically different ways: the NCBI might use BLAST, UCSC might use BLAT, and the EBI the SSAHA search engine.

For bench researchers, the web-services world might not look very different from the current one. Online databases would still exist, each with its own distinctive character and user interface. But bioinformaticists would now be able to troll the online databases to aggregate data simply and reliably. Furthermore, software engineers could create standard user interfaces that would work with any number of online data sources. This opens the door to genome 'portals' for those researchers who prefer to access multiple data sources from a single familiar environment.

A challenge for web services is to deter-

Box 1: Data provider's code of conduct

The downside of the web-services world is that it won't be achieved overnight. It will be some years before web services are stable and complete enough to replace current sites. Meanwhile, I propose the following bioinformatics data provider's code of conduct, to maximize the usefulness and re-usability of a data source.

1. A web page is an interface

Although web pages were designed for access by people, they can be accessed by scripts. Guide this trend, don't fight it. Online databases should provide the rules for linking to pages, including the hours and frequency available for scripts to download information.

2. An interface is a contract

Once a web page is in use by screen scrapers, it becomes a contract between the data provider and the data consumer. Changing the interface violates the contract. Document each interface and warn if it is unstable. If an interface needs to change, provide users with plenty of advance warning. When a widely used interface is changed, give it a new URL and maintain the legacy interface for a while.

3. Choice is good

Make your information available in several formats for bioinformaticists with different needs and abilities. HTML is the least desirable format to publish data for use with bioinformatics scripts. Tab-delimited text is easily handled by scripts and is suitable for many types of simple data. XML is harder to parse but can convey much more complex information. A SOAP/XML interface may not be accessible to novice bioinformaticists, but will be greatly appreciated by the more advanced ones.

4. Allow batch downloads

Make the whole data set available for batch download, breaking it up into logical, bite-sized pieces if necessary. Many developers have found themselves writing scripts to download entire databases one web page at a time as the only route to a full data set, which is wildly inefficient.

5. Use existing file formats

Avoid reinventing wheels: use existing file formats when possible. For example, there are already enough formats to describe features on a sequence (GenBank, EMBL, GFF, BSML, Agave, GAME, DAS), so it is doubtful that the world needs another. There are also good formats to describe genome assemblies, microarray experiments and results, three-dimensional structures and bibliographic references.

6. Design sensible formats

If an existing format doesn't support the information that a data source wants to publish, use common sense in designing new formats. Use tab-delimited text if it is sufficient. If the data are hierarchical, XML is a natural choice. It is better to start simple than to create a complex format to cover all contingencies.

7. Allow ad hoc queries

Support a true query language. Researchers often search for hidden relationships among biological data, but browsing through a set of hyperlinked pages may not be the best way. Many online databases have created web-based forms that generate summary reports based on simple filters. Better still would be to make copies of your databases available for direct access using the database's native query language. Although this is a significant investment, it is well worth the effort.

mine what data sources implement which services. The solution is a service registry. To advertise its services, a data source registers its services with a designated directory server on the Internet. Then, when a script needs access to a particular service, it consults the registry to find out what data sources provide the service. If the same service is provided by more than one data source, the script consults the user for help, or uses built-in rules to choose the data source, then returns the results.

Although this proposal may seem a far cry from what happens now, the technology exists to make it a reality. The World Wide Web Consortium, with industry heavyweights such as IBM and Microsoft, is promoting an alphabet soup of standards: SOAP/XML, WSDL, UDDI and XSDL. These are already being used by some data providers: examples include those sites that publish genomic annotation data using the distributed annotation system format (www.biodas.org) and the EBI's SOAP-based bibliographic query service (industry.ebi.ac.uk/openBQS). Also being developed is a highly promising integration platform called Omni-

gene (omnigene.sourceforge.net); Integr8, an ambitious integration project run by the EBI (www.ebi.ac.uk); and a pharmaceutical industry-backed project called I3C (www.i3c.org). Finally, the caBIO project (ncicb.nci.nih.gov/NCICB/core/caBIO) at the US National Cancer Institute (NCI) is a comprehensive open-source project that aims to make the NCI's cancer databases available via a web services architecture.

The risk, of course, is that like the mediaeval Italian city-states, each of these projects will endorse its own idea of standardization, and a chaotic world of incompatible bioinformatics data standards will be replaced by a chaotic world of incompatible web-service standards. We can look forward to a bit of a struggle before one set of standards achieves pre-eminence, but I have no doubt that unity will be reached eventually. ■

Lincoln Stein is at the Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. This Commentary is adapted from a keynote speech given by the author at the 2002 O'Reilly Open Bioinformatics Conference in Tucson, Arizona, USA.

Confirming a bold prediction

A look back at work to prove the semi-conservative replication of DNA.

Meselson, Stahl, and the Replication of DNA: A History of "The Most Beautiful Experiment in Biology"

by Frederic Lawrence Holmes

Yale University Press: 2001. 416 pp. \$40

Robert Olby

So much attention is given to DNA today that no small effort is required to recapture the climate of uncertainty and caution that met the launch of the double-helix model of its structure by James Watson and Francis Crick in 1953. Looking back in the light of more recent advances in molecular biology, that year seems to mark the decisive turning point. If so, why, one might wonder, did James Watson worry so much that the double-helical structure might prove to be wrong? And why did he advise his co-discoverer against accepting an invitation in 1953 to speak about their structure on the radio?

Consider, too, the research of two young postdocs — Matthew Meselson and Frank Stahl — in the years 1957–58, as they worked to find evidence for the semi-conservative form of DNA replication. Was this work simply an elegant experiment confirming a prediction of a model made five years before? Granted, their experiment appears to be both beautiful and compelling, but did we really need it? Holmes, a seasoned historian and professor at Yale University, leaves us in no doubt that we did. He treats us to a thorough analysis of the context and origin of this famous experiment, the lives of its two authors during the work, the impact of the published results, and finally a soliloquy on the nature and role of experiment in molecular biology.

The standard account is by now familiar. Watson and Crick's model for DNA suggested a molecular mechanism for gene replication. The chains of the duplex would separate, each then providing a template upon which a complementary chain could form, thus generating two 'daughter' double helices. According to the model, each daughter double helix carries one of the parental chains. The replication, as Gunter Stent put it, is 'semi-conservative'. There was a problem, though: how to unwind the entwined helices in the first place. This was a formidable task but not impossible, judged Crick, but to Max Delbrück it seemed insuperable. Caught between the scepticism of Delbrück and the confidence of Crick, Watson trod a fine line, motivated by fear that the structure might be wrong and hope that it would turn out to be right.

Mercifully, along came Meselson and



Pairing up: Frank Stahl (left) and Matthew Meselson nailed the mechanism of DNA replication.

Stahl with their wonderful experiment that demonstrated the semi-conservative nature of the replication process in the bacterium *Escherichia coli*. Published in 1958, it offered a stunning verification of the original prediction. Meselson and Jean Weigle extended the result to phage- λ in 1959. This was followed by John Cairns' demonstration by autoradiography of the molecular unity of the phage- λ chromosome in 1962, which clinched the case for identifying the subunits as single chains.

Reading this book should disabuse us of such easy impressions of life in the early days of molecular biology. Holmes has mined the archives, probed the scientists' memories, and absorbed the abundant literature, both the great and the not-so-great, and that which in retrospect we can judge the 'wrong' and the 'right'. The result is one of the most authentic narratives in the history of molecular biology — not a 'winner-takes-all' story, but a clash of differing views, a striving and a resolution. When the manuscript was finished, Meselson read it and remarked: "It brought it all back. I can remember being in that world — it's imperfect, but a time machine."

In the onward march of science, things that are resolved are 'black-boxed' — students are no longer taught how they were established — and no textbook of molecular biology today would include a diagram of the 'dispersive' replication scheme preferred by Delbrück, in which there are breaks and

rejoinings in the chains every five bases to overcome the unwinding problem. Holmes points out that even the Meselson–Stahl experiment seems to be disappearing from the newest texts. Here he has recaptured the technical and intellectual context of the experiment, and explained why it was not judged by its authors to constitute a demonstration of the semi-conservative replication of the DNA molecule, but only, as they put it in 1958, of 'DNA subunits'. They were clear, however, that they had destroyed the two alternative suggestions: conservative and dispersive replication.

Using the analytical ultracentrifuge at Caltech run by Jerome Vinograd, and their innovative caesium chloride gradient, Meselson and Stahl were able to distinguish three distinct bands representing DNA: one 'light' band from bacteria grown in a normal medium, one 'heavy' band from bacteria nourished with an isotopic 'heavy' nitrogen source, and, in between, one 'hybrid' band representing DNA from the progeny of the latter formed under normal 'light' nutrition. These progeny DNA samples, they inferred, were composed of equal quantities of heavy and light subunits — one from the parental 'heavy' DNA, the other 'light' and freshly minted. The experiment had thus a visual beauty to add to its other merits: the apparently clean separation of the bacterial growth on heavy and light media, and the strategy of the experiment that allowed it to meet possible objections,

so it even swayed the reluctant Delbrück. This was no small achievement, to be put alongside the simplicity and substantial benefit to molecular biology of caesium chloride gradients for ultracentrifugation.

Holmes suggests that the growing knowledge of the immense complexity of the mechanism required for DNA duplication may be starting to crowd out mention of the Meselson–Stahl experiment. When Watson and Crick suggested in 1953 that DNA replication might occur without the aid of a special enzyme, little did they foresee the complexity of the process, involving as it does not only DNA polymerase, but also the DNA helicases and primases, the proof-reading exonuclease, the sliding DNA clamp and the Okasaki fragments — a multi-enzyme system that performs its task with remarkable fidelity. Nor, perhaps, did they expect the unwinding problem to resurface again to challenge the double helix as late as the 1970s.

This book joins a growing library of outstanding historical studies of modern experimental biology that includes Hans-Jörg Rheinberger's study of protein synthesis in the cell-free system, *Toward a History of Epistemic Things* (Stanford University Press, 1997) and Nicolas Rasmussen's analysis of the adaptation of the electron microscope for the study of the cell, *Picture Control* (Stanford University Press, 1999). Their scholarship is impeccable and their analyses perceptive and revealing, not just of an age fast receding into the past, but of what Bruno Latour called "science in the making". ■

Robert Olby is in the Department of the History and Philosophy of Science, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA.

When ejaculates collide...

Sperm Competition and its Evolutionary Consequences in the Insects

by Leigh W. Simmons

Princeton University Press: 2001. 434 pp. £24.95 (pbk)

Michael T. Siva-Jothy

A million million spermatozoa

All of them alive

Out of their cataclysm but one poor Noah

Dare hope to survive.

And among that billion minus one

Might have chanced to be

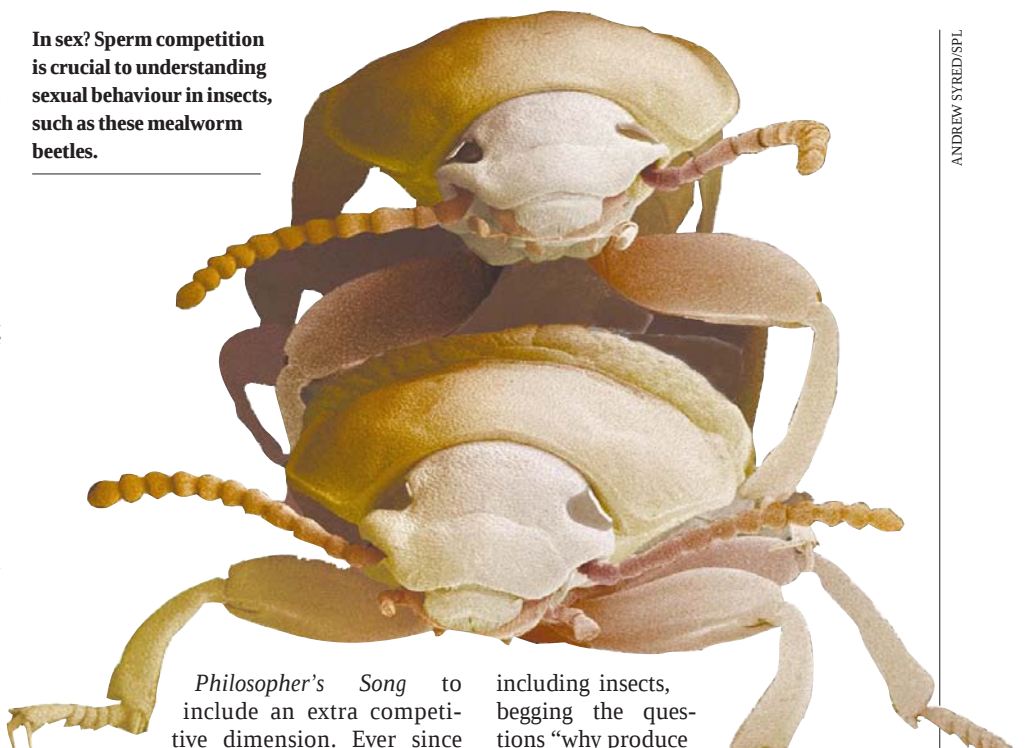
Shakespeare, another Newton, a new Donne

—

But that One was Me.

If Aldous Huxley had lived another seven years he might have amended his *Fifth*

In sex? Sperm competition is crucial to understanding sexual behaviour in insects, such as these mealworm beetles.



Philosopher's Song to include an extra competitive dimension. Ever since

Antoni van Leeuwenhoek first looked at sperm under a microscope, and probably long before, humankind has had an inkling of the wastage that occurs in male gametes.

But it was not until 1970 that a formal theoretical framework, and some elegant empirical support for it, emerged. Geoff Parker realized that when two or more males mated with the same female, selection would favour the evolution of male traits that gave one sperm-owner a paternity advantage. This elegant and, with the benefit of hindsight, obvious concept opened the floodgates to a sea of ideas and explanations. Suddenly post-copulatory guarding behaviour made sense, we began to understand the causal basis of variation in hitherto ignored aspects of reproductive behaviour, and we had an elegant explanation for a million million spermatozoa.

Parker's theory had a profound effect on the fledgling field of behavioural ecology — he laid a cornerstone that defined it and illustrated the value of combining modelling, empiricism and a productive disregard for disciplinary boundaries to address questions about the how and why of evolution. Parker used insects as his model system: they were particularly susceptible to this form of selection and were amenable to the combined physiological, anatomical and behavioural studies that revealed the nature of the adaptations that resulted.

Having served his intellectual apprenticeship in Parker's laboratory and played a significant part in directing the field, Leigh Simmons is better qualified than most to define the current state of sperm competition. However, there are already three books that examine sperm competition in a range of taxa,

including insects, begging the questions "why produce another?" and "why just insects?" Simmons argues that a comprehensive review of our current knowledge of the phenomenon in insects provides the broadest platform on which to base any understanding of sperm competition. To make this point he picks the field apart, paper by paper, and skilfully reassembles it around the important conceptual issues, thereby bringing sperm competition, rather than insect behavioural ecology, into crisp focus.

Each chapter starts, when appropriate, with the theoretical basis for its sub-theme, followed by a well-structured review of the empirical work. For example, the chapter on behavioural adaptations to avoid sperm competition begins with a very readable distillate of the theory and modelling of mate-guarding. (Similar summaries of the critical modelling and theory preface other chapters and are a real bonus.) The chapter then reviews the substantial literature on mate-guarding and alternative hypotheses for this behaviour, before moving on to the more contentious subject of male mate choice. By leaving the open-ended subjects until the end of their chapters, Simmons contextualizes the new areas and so guides the reader to paths that have potential for future work.

This book is an excellent summary of a recent, important and relatively large addition to our understanding of evolution and its consequences. It is an essential read for anyone interested in the reproductive determinants of fitness. ■

Michael T. Siva-Jothy is in the Department of Animal and Plant Science, University of Sheffield, Sheffield S10 2TN, UK.

An increase in insanity

The Invisible Plague: The Rise of Mental Illness from 1750 to the Present

by E. Fuller Torrey & Judy Miller
Rutgers University Press: 2001. 400 pp. \$28
Hugh Freeman

That the 'strain of modern life' causes more mental illness is a refrain that goes back several centuries at least. There is certainly evidence for some increase in the prevalence of depression in recent decades, at least in industrialized countries. But for the most serious of mental disorders—schizophrenia—trends are, if anything, now in the other direction. In the view of Fuller Torrey and his colleague Judy Miller, the 'modern life' that may have provoked the rise in mental illness was in fact some two centuries ago.

Torrey is a distinguished US researcher in psychiatry, a prolific writer for both scientists and the public, and director of the Stanley Foundation. His extensive work on schizophrenia—notably using twins—has led him to be increasingly critical of conventional views that the frequency of the illness does not vary much worldwide and has not varied over time. Its presence in practically every known human society does seem to have been confirmed, but identifying it as a relatively 'new' disease is more controversial.

The first scientific suggestion of this kind was by the British psychiatrist Edward Hare in the early 1980s, but acknowledgement

here of this work is rather sparse and only one of his two important papers is referenced. The crux of his argument is that there was a massive but steady rise in the number of patients with 'lunacy' in mental hospitals from the early nineteenth century to the mid-twentieth. Of course, for there to be in-patients, there have to be hospitals, so the story is also one of building a huge system of asylums in every country that could afford it. In *The Invisible Plague*, Torrey and Miller draw evidence from England, Ireland, the Maritime provinces of Canada and the United States; Scotland is also included but has a rather marginal position.

Useful information about this question first appears in English in the late sixteenth century, when madness was clearly recognized. There are 20 references to it by Shakespeare alone, but Torrey and Miller believe that it remained relatively rare for the next 200 years. Insanity seems to have been regarded then mostly as a temporary state, secondary to some medical condition, rather than as a disorder in its own right. In the late eighteenth century there was increasing public concern in England about ill-treatment of the insane, as part of a wider humanitarian movement. Its most important expression was the foundation by the Quakers in 1792 of the York Retreat, where 'moral treatment' replaced chains and brutality. The authors say that the first clear description of schizophrenia appeared four years later.

As the nineteenth century began, the first public asylums were established in Britain, and then it was uphill all the way, both in the number of hospitals and of patients. That

the number of the institutionalized insane increased far more rapidly than the general population is clearly documented: in all four of the countries concerned, from the mid-eighteenth to the mid-twentieth century the number of institutionalized insane as a proportion of the general population increased sevenfold. The highest figure ever in England or Wales—almost 4 per thousand population—was reached in 1915. Why this was the peak is not entirely clear, but in some other countries the growth in numbers continued with the building of new hospitals, a process that had practically finished in Britain.

In these calculations, 'insanity' has been taken as roughly equivalent to schizophrenia, although the Victorian asylums contained many patients with tertiary syphilis, alcoholism and senile dementia. In 1870 there were also 12,000 'lunatics' among the poor in English workhouses, and it was not until the 1950s that provision for the poor was clearly separated from that for the sick. The argument put forward in the book, though, is that such considerations make little difference to the concept of a huge increase then in the number of institutionalized insane in all the countries concerned.

This view of schizophrenia as a new and epidemic disease has not gone unchallenged. Historians following the philosopher Michel Foucault see the rise of asylums as the removal of troublesome people from general society, although examination of how the early institutions were set up shows that humanitarian concern was the strongest motive. Alternative views to the epidemic theory have emphasized the effects of migration, better medical care (so that asylum patients lived longer), urbanization (which made the insane more visible) and diminished tolerance by larger communities. But Torrey and Miller believe that no such factors can account for the increase in the identified insane.

If there was an epidemic, though, it must have had a cause. The authors argue that "novel approaches" are needed to investigate what this was, but mention only diet, alcohol, toxins and infections as candidates—none with much conviction. This epidemic, they say, has been "so insidious that after 200 years, it has barely been noticed. It affects over four million Americans, but is not recognised as a major public health problem."

This highly informative and stimulating work has certainly raised some neglected questions that demand more serious scientific attention. For one thing, the burgeoning cities of developing countries might be a fertile soil for schizophrenia, although there is no evidence yet that this is happening. ■

Hugh Freeman of 21 Montagu Square, London W1H 2LF, UK, is an Honorary Visiting Fellow at Green College, Oxford. He is former editor of the British Journal of Psychiatry.



Head count: the number of institutionalized insane increased rapidly in the nineteenth century.

The jet set

Mario Livio

Although a wide variety of astrophysical objects produce powerful jets, we still lack a comprehensive theory of their formation. In our Galaxy, young stellar objects, massive X-ray binaries, black hole X-ray transients, symbiotic stars, supersoft X-ray sources and even some planetary nebulae are all accreting systems that produce jets. Among extragalactic sources, powerful jets are observed in active galactic nuclei (accreting, supermassive black holes) and are thought to exist in γ -ray bursts (which possibly involve accreting, stellar-mass black holes).

Despite the ubiquity of jets, there is no generally accepted theory for the mechanism of their acceleration and alignment. Early attempts to explain this collimation involved nozzles (similar to rocket exhausts), in which an adiabatic flow propagating in a medium with decreasing pressure first converges and then diverges as it accelerates to supersonic speed. Another idea was that the collimating agents are funnels that are formed at the centres of dense tori. Radiation pressure on electron-positron pairs, on resonance lines,

or on dust, have been suggested to accelerate jets in some of the systems mentioned above.

The main problem of these early models is that they do not work for all classes of astrophysical jets. For example, although the power emitted by the central source approaches the critical Eddington luminosity (at which radiation pressure exactly balances gravity) in supersoft X-ray sources, this is not true for most other classes of objects. The drag caused by radiation limits the attainable speeds to values below those observed in other (ultrarelativistic) jets. Similarly, dense tori with narrow axisymmetric funnels, which were once assumed to exist around black holes, were later shown to be unstable to non-axisymmetric instabilities that could destroy them on a dynamic timescale. Can a universal mechanism of acceleration and collimation that operates in all classes be found?

The signs are positive. First, there is strong observational evidence that almost all jet-producing systems contain an 'accretion disc' around a compact object. This disc is both a source of energy (in the compact object's gravitational potential) and provides the required axial symmetry. Second, jet velocity is always of the order of the escape velocity from the central object, indicating that jets originate from the centre of the accretion disc. The Galactic black hole transient source GRS 1915 + 105 (a microquasar) provides even more evidence for the connection between the jet and the central part of the disc, and we can begin to see a picture in which the inner disc is episodically accreted while ejecting a relativistic plasma, which subsequently produces infrared and radio flares by synchrotron emission.

The most promising universal mechanism for jet acceleration and collimation relies on an accretion disc threaded by a poloidal, large-scale magnetic field. The basic idea is that some magnetic flux is in open field lines, which form a certain angle with the disc's surface. Ionized material is forced to follow field lines; as these lines are anchored in the disc and rotate with it, material is centrifugally accelerated like a bead on a wire. The jet may be collimated by hoop stresses resulting from the magnetic field's toroidal component; alternatively, if the poloidal magnetic flux is largest at the disc's edge, and the disc's radius is large relative to that of its central object, the jet will be collimated by the magnetic field lines.

Some differences in behaviour remain unexplained by a unified mechanism. For example, jets from the nuclei of 'blazars' — active galaxies from which jets are beamed in

Astrophysical spouts

Despite the ubiquity of astrophysical jets, there is no generally accepted theory for the mechanism of their formation.

the direction of the observer — are ultra-relativistic, whereas jets from Seyfert galaxies are not. Yet in both cases, jets are thought to originate from the centres of discs around black holes. The difference may be due to different environments, different mass-loading of the magnetic field lines, or something else altogether.

Extraction of rotational energy from a rapidly spinning (Kerr) black hole by magnetic fields that permeate the hole's event horizon is often invoked as a means of launching jets in active galactic nuclei. The mechanism itself can be understood if the black hole is likened to a resistor rotating in a magnetic field and generating a potential difference between the hole's pole and equator. X-ray observations of the Seyfert galaxy MCG-6-30-15 and the Galactic black hole candidate XTE J1650-500 have now revealed an extremely broad and redshifted iron $K\alpha$ line, indicating an origin in the very centre of the accretion disc. This implied emissivity may mean that energy is indeed extracted from the spinning black hole. If true, this could show that an exotic process anchored in electromagnetism in curved spacetime is operating across a range of a factor of a million in black-hole mass!

This interpretation, however, is far from unique. Much better understanding of processes within the plunging orbits between the innermost stable orbit in the accretion disc and the black hole's event horizon will be needed to draw a final conclusion. ■

Mario Livio is at the Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA.

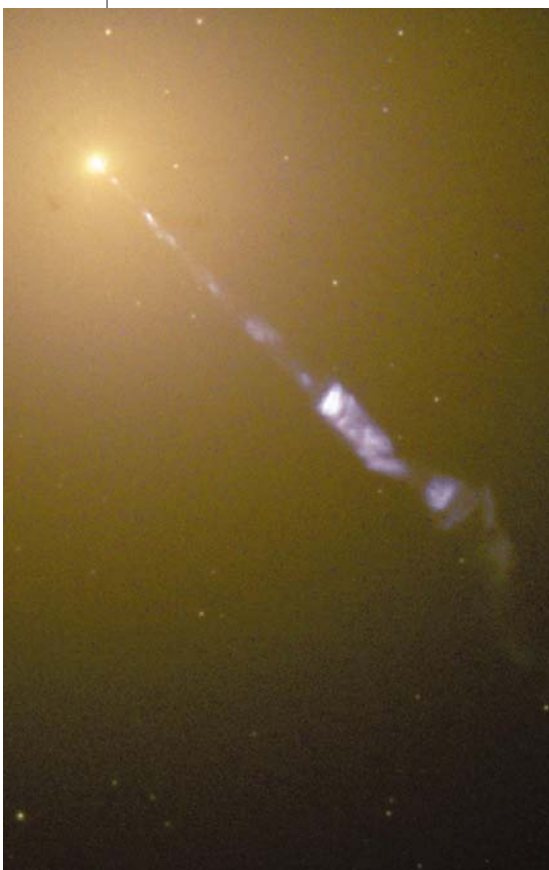
FURTHER READING

Mirabel, I. F. & Rodríguez, L. F. *Nature* **392**, 673–676 (1998).

Wilms, J. *et al.* *Mon. Notices R. Astron. Soc.* **328**, L27–L31 (2001).

Celotti, A. & Blandford, R. D. in *Black Holes in Binaries and Galactic Nuclei: Diagnostics, Demography and Formation* (eds Kaper, L. *et al.*) 206–215 (Springer, Berlin, 2001).

Livio, M. in *Probing the Physics of Active Galactic Nuclei by Multiwavelength Monitoring* (eds Peterson, B. M. *et al.*) 225–248 (ASP, San Francisco, 2001).



Shining example: the M87 galaxy spews out its jet.

A lowdown on oxygen

John M. Hayes

We can hope that the term 'palaeophosphatometry' won't catch on. The approach it describes has nonetheless delivered a plausible — if partial — answer to one of the main questions about Earth's history.

There are two big questions about the early history of oxygen-producing — oxygenic — photosynthesis. When did it start? And why did it apparently then take so long to push atmospheric concentrations of oxygen to their present levels? The first question has a specific answer, yet to be pinned down precisely. The second question probably has several answers, relating to processes rather than to any point in time. And they must be fundamental, to say the least. On page 159 of this issue¹, Bjerrum and Canfield give us something fundamental to think about. Specifically, they provide new evidence for low concentrations of dissolved phosphate in Earth's early oceans. By limiting the growth of algae, low phosphate levels would have significantly weakened the biological source of O₂.

Dates nominated for the beginning of oxygenic photosynthesis range from 3.5 billion to 2.75 billion years ago. The earlier limit is based on an interpretation of microfossil forms and the textures of stromatolites — sedimentary structures that have been interpreted as cyanobacterial mats dating to that time (see ref. 2 for review). Cyanobacteria are microorganisms that are capable of oxygenic photosynthesis, and are thought to have been the main producers of oxygen for much of Earth's history. However, the 3.5-billion-year-old deposits are being reassessed, and the view that they contain cyanobacteria is not likely to survive^{3,4}. The later date was based originally on an interpretation of the carbon-isotope record⁵ but has since been buttressed by the identification, in related sediments, of molecular-carbon skeletons that are characteristic of cyanobacteria and even more complex oxygenic photosynthetic organisms⁶. So, although the earlier date may yet prove to be correct, 2.75 billion years is probably a reliable minimum age for the advent of oxygenic photosynthesis.

By most interpretations, it then took well over two billion years for concentrations of oxygen in the atmosphere to reach modern levels. The most profound indicator of when these levels were reached is the evolution of animals with complicated body plans that require the diffusion of oxygen across several layers of cells. Such animals appeared only around 600 million years ago, and the fact that they did not evolve earlier is often attributed to low levels of oxygen⁷. There is also plenty of geochemical evidence that oxygen

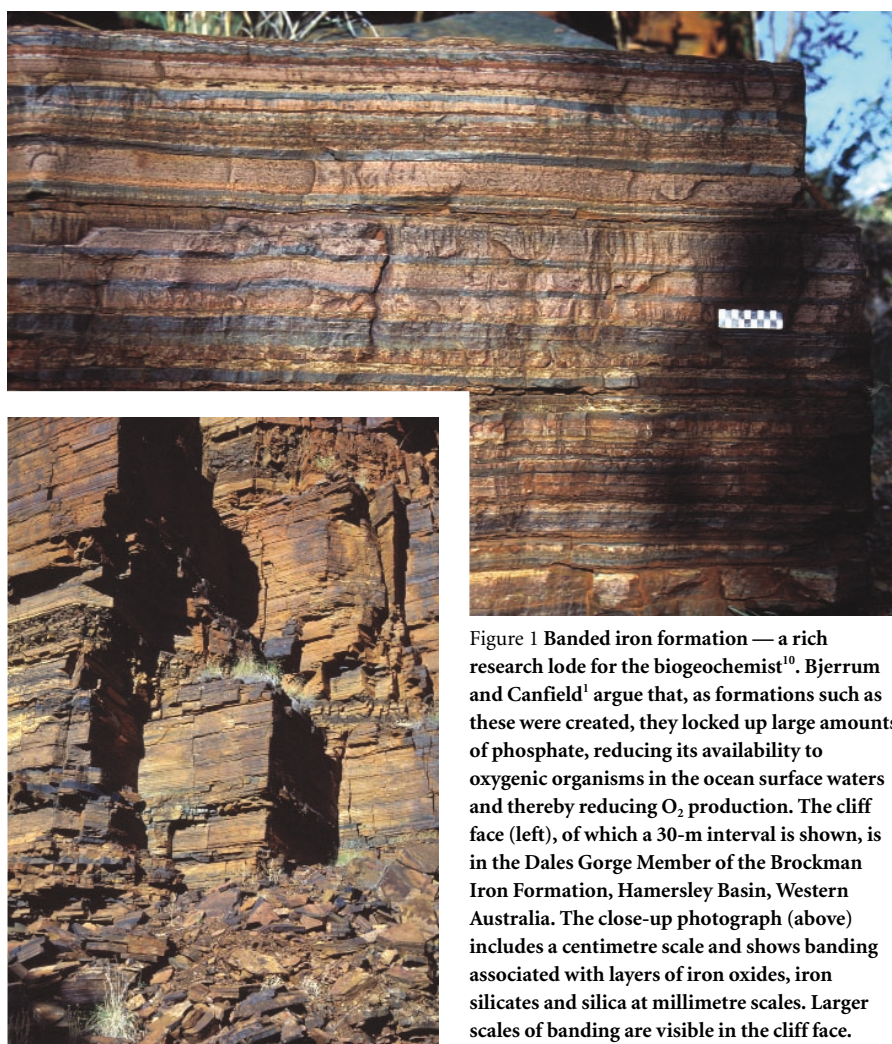


Figure 1 Banded iron formation — a rich research lode for the biogeochemist¹⁰. Bjerrum and Canfield¹ argue that, as formations such as these were created, they locked up large amounts of phosphate, reducing its availability to oxygenic organisms in the ocean surface waters and thereby reducing O₂ production. The cliff face (left), of which a 30-m interval is shown, is in the Dales Gorge Member of the Brockman Iron Formation, Hamersley Basin, Western Australia. The close-up photograph (above) includes a centimetre scale and shows banding associated with layers of iron oxides, iron silicates and silica at millimetre scales. Larger scales of banding are visible in the cliff face.

levels were indeed low before that time. In other work, Canfield⁸ has shown that deep ocean waters probably contained persistent traces of hydrogen sulphide between 1.8 billion and 0.6 billion years ago. If this is true, it shows that oxygen concentrations in surface waters (and in the atmosphere) were not high enough to drive O₂ into deeper waters, where it would spontaneously oxidize the hydrogen sulphide.

Still earlier, more than 1.8 billion years ago, the sediment record is dotted with banded iron formations (Fig. 1). These are massive deposits that formed when iron oxides, carbonates, silicates and sulphides settled to the sea floor, much as calcium carbonate and silica do now. As sites of focused deposition, they show clearly that iron was

mobile in the deep sea. Mobility requires solubility. This, in turn, requires the absence of O₂. Iron in the ferric (trivalent) oxidation state, which would be formed if O₂ were present, precipitates rapidly at pH values consistent with the formation of limestones, which are widespread from these same times. The resulting picture is one in which oxygen producers existed in surface environments but somehow failed to oxygenate the Earth's atmosphere and oceans to modern levels.

So what might have been happening? Was less oxygen being produced by the photosynthetic sources? Or was more being sucked up into 'sinks' of some type or another? Most previous thinking has focused on sinks. As oxygen was produced, it would have been consumed by the unoxidized iron and

sulphur being delivered to the ocean by weathering and erosion of the earliest rocks. It would have taken time to saturate this source of unoxidized material — at least as long as it takes for several global cycles of erosion and deposition to take place (something like 600 million years). But the records are not good enough, either in terms of the total inventories or of the rates of delivery of iron and sulphur, to develop a satisfyingly quantitative view of the dynamic role of sinks.

In 1984, Holland⁹ provided the clearest outline to date of the geochemical dynamics of oxygen on the early Earth. In doing so, he stressed that levels of O₂ would be controlled dynamically by balances between sources and sinks, and he raised the specific possibility that the strength of the source might be limited: "It is instructive to see how the marine geochemistry of phosphorus might have exercised a dominant control on biological productivity and on the oxygen content of the Precambrian atmosphere." The logic is straightforward: the production of O₂ requires cells; cells require nucleic acids and phospholipids; these compounds cannot be synthesized, and O₂ cannot be produced, without a supply of phosphate; limiting that supply can therefore limit the production of O₂.

It is this idea that Bjerrum and Canfield¹ have now quite beautifully extended. To estimate the abundance of phosphate in ancient oceans, they have taken another look at the banded iron formations, not only viewing them as sinks for O₂, but also pointing to the concentrations of phosphate that they contain. The surfaces of iron-rich minerals adsorb phosphate strongly, and this adsorption can be described in terms of a distribution coefficient. Given the value of this coefficient (determined from observations of modern iron-rich precipitates), Bjerrum and Canfield have estimated the concentration of phosphate in ancient sea water. The answer turns out to be 0.15–0.6 micromolar, very much lower than the modern value of 2.3 micromolar. By considering likely balances, the authors show that this low concentration could be due to the precipitation of the banded iron formations, which would have created a huge sink for phosphate. As the phosphate was adsorbed, carried to the sea floor and buried under accumulating sediments, it would not be available as a nutrient for cyanobacteria.

To calibrate their 'phosphatometer' and obtain a relationship between phosphate in sea water and in iron-rich precipitates, Bjerrum and Canfield have referred to studies of hydrothermal precipitates forming in the modern ocean. It is not certain that these precipitates accurately represent the particles that were present in the ancient oceans, nor is it certain that current concentrations of phosphate in iron formations faithfully represent those that were initially present. But it is unreasonable to ask for certainty about such

details at a distance of two billion years or more. The new work makes a strong case that concentrations of phosphate in the ocean between 3.2 billion and 1.8 billion years ago were distinctly lower than at present.

The authors themselves raise a further question. Deposition of banded iron formations ceased about 1.8 billion years ago. What happened to the balance between oxygen sources and sinks then? If phosphate had been the limiting nutrient and its concentrations increased, did some other nutrient just happen to step into the limiting role? The most obvious candidates are nitrate and, given the possibility that metal ions were swept from deep ocean waters by precipitation as sulphides, iron itself.

John M. Hayes is in the Department of Geology and

Geophysics, Woods Hole Oceanographic Institution, 360 Woods Hole Road, Woods Hole, Massachusetts 02543-1539, USA.

e-mail: jhayes@whoi.edu

1. Bjerrum, C. J. & Canfield, D. E. *Nature* **417**, 159–162 (2002).
2. Buick, R. in *Palaeobiology II* (eds Briggs, D. E. G. & Crowther, P. R.) 13–21 (Blackwell Science, Oxford, 2001).
3. Schopf, J. W. *et al.* *Nature* **416**, 73–76 (2002).
4. Brasier, M. D. *et al.* *Nature* **416**, 76–81 (2002).
5. Hayes, J. M. in *Early Life on Earth: Nobel Symposium 84* (ed. Bengtson, S.) 220–236 (Columbia Univ. Press, New York, 1994).
6. Brooks, J. J., Logan, G. A., Buick, R. & Summons, R. E. *Science* **285**, 1033–1036 (1999).
7. Runnegar, B. *Palaeogeogr. Palaeoclimatol. Palaeoecol. (Glob. Planet. Change)* **97**, 97–111 (1991).
8. Canfield, D. E. *Nature* **396**, 450–453 (1998).
9. Holland, H. D. *The Chemical Evolution of the Atmosphere and Oceans* 413–420 (Princeton Univ. Press, New Haven, Connecticut, 1984).
10. Klein, C. & Beukes, N. J. in *Proterozoic Crustal Evolution* (ed. Condie, K. C.) 383–418 (Elsevier, Amsterdam, 1993).

Cytoskeleton

Microtubules do the twist

Patrick J. Hussey

As plant life diversified during evolution there would have been intense competition for light, making the ability to twist and climb advantageous. A subcellular filamentous network can create the twist seen in climbing plants.

The love affair between a left-handed-twisting plant, bindweed, entwining the right-handed-twisting honeysuckle was doomed when a bee told them to consider their offspring:

Poor little sucker, how will it learn
When it is climbing, which way to turn?
Right — left — what a disgrace!
Or it may go straight up and fall flat on its face!*

And the honeysuckle and bindweed are engraved in the minds of botanists as well as music lovers. There are numerous twisting or twining plants; the shoots of some turn to the right (clockwise) while others twist to the left (anticlockwise). External conditions such as light, heat or humidity cannot influence this asymmetry. As Austrian botanist Anton Kerner von Marilaun wrote¹, the shoot "continues to twist and twine according to an innate tendency inherited from generation to generation, and we can only refer the different directions of twisting to internal causes, to the peculiar constitution of the living protoplasm in each particular plant".

On page 193 of this issue, Thitamadee and colleagues² put these observations on a molecular footing. The thale cress *Arabidopsis thaliana* is usually 'straight', but the authors have identified mutant *Arabidopsis* plants with shoots and roots that twist to the left. They have called the mutation *lefty*, and

found that it compensates for the right-handed twist of the previously known *Arabidopsis* mutation *spiral1* — so an anticlockwise plus a clockwise twist equals a straight plant. The authors tracked down the mutation in *lefty* plants specifically to one of the two major proteins that make up subcellular filaments called microtubules.

In a typical root, files of cells extending away from the root tip are produced by the successive division of precursor cells in a structure known as the meristem, near the tip. In normal *Arabidopsis* plants, cell files in the root epidermis (the outer layer of cells) are straight, but, as Thitamadee *et al.*² show, in the *lefty* mutants these files are twisted anticlockwise. In normal plants the microtubules in the cell periphery form a random pattern in meristem cells; in the elongated cells further from the root tip microtubules are arranged at right angles to the long axis of the cell. By contrast, in the *lefty* mutants² the microtubules in the elongated epidermal cells are offset at a different angle to the long axis, and form right-handed helical arrays.

It has long been thought that microtubules guide the deposition of the fibres that constitute the plant cell wall, and that the fibres lie parallel to the microtubules (although there are exceptions). The cell wall maintains the cell's shape, which is determined by the underlying microtubules. Offsetting the angle of the microtubules to form right-handed helices may offset the orientation of the fibres that make up the cell wall, resulting in a twist that is propagated along a cell file. This is probably the underlying

*From Flanders and Swann's song *Misalliance*, reproduced by permission of the Estates of Michael Flanders and Donald Swann 2002.

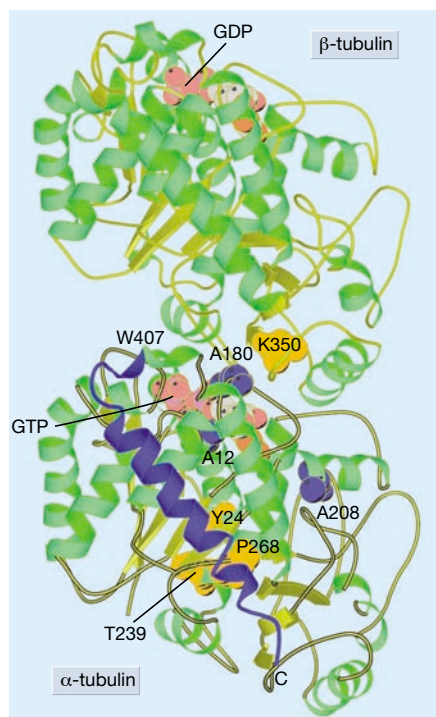


Figure 1 A basis for twistiness? This structure of the α -tubulin/ β -tubulin dimer shows the positions of amino acids that are known to be mutated in several plants or algae. Purple, mutations that affect twisting. The mutation of alanine (A) — or serine — to phenylalanine at position 180 in α -tubulin causes the *lefty* plants identified by Thitamadee *et al.*². Mutation of alanine 12 to valine, or of alanine 208 to threonine, or deletion of tryptophan (W) 407, suppresses the left-handed twisting of *lefty* plants. Orange, mutations that cause plant resistance to dinitroaniline herbicides. (Y, tyrosine; P, proline; T, threonine; K, lysine.) Kindly provided by L. Amos, using coordinates from ref. 7.

cause of the left-handed twisting of *lefty* plants. Similarly, microtubules in cells of the clockwise-twisting *spiral1* mutant plants form left-handed helices³.

But Thitamadee *et al.*² delve deeper still, and find the precise protein affected by the *lefty* mutation: the α -subunit of the major microtubule protein tubulin. (The protein affected by the *spiral1* mutation is not known but might also be part of the microtubule network³.) Tubulin is a dimer of α - and β -subunits and is dumb-bell-shaped; many dimers associate head-to-tail to form protofilaments. Generally, 13 protofilaments associate laterally in a staggered fashion to form a hollow tube — the microtubule — with an external diameter of 25 nanometres.

Thitamadee *et al.* find that the *lefty* mutation results in an amino-acid change in α -tubulin: the serine at position 180 in the amino-acid chain is replaced by phenylalanine. The amino acid at this position is obviously important as it is highly conserved, being either serine or alanine, across taxonomically distinct species. It is close to the

interface between the α - and β -tubulin subunits and is in contact with the nucleotide guanine triphosphate (GTP) (Fig. 1). The function of this GTP is unclear; unlike many such nucleotides, it is not hydrolysed to release energy and then replaced by another GTP. But it is possible that its presence favours the formation of the tubulin dimer.

So, mutations at or near this site might change the shape of the dimer, resulting in less stable microtubules or a change in microtubule shape. It would be interesting to see electron micrographs of microtubules from *lefty* plants. It is known that certain mutations in the FtsZ protein (the bacterial equivalent of tubulin) cause it to form a spiral instead of the normal closed ring and that this causes morphological abnormalities⁴. Perhaps the same happens in *Arabidopsis*, too. Whatever the case, though, the net result is that the microtubules in elongated cells of *lefty* plants are offset from the normal transverse orientation. Interestingly, the effects of the *lefty* mutation are overcome by mutations at other sites in α -tubulin². Some of these (Fig. 1), which are near the α -tubulin/ β -tubulin interface (changing alanine at position 208 to threonine) or near GTP (changing alanine at position 12 to valine), may 'tweak' the dimer back into its normal shape, and result in normal, straight *Arabidopsis* plants.

Thitamadee *et al.*'s results² suggest that microtubules can do the twist, and could be Kerner's "peculiar constitution of the living protoplasm". A perhaps confounding fact is that normal left- and right-twining plants, as well as the upright *Arabidopsis*, probably all

have serine or alanine at position 180 in α -tubulin. So this may not be the precise amino acid that is different between twisting and non-twisting plants. But amino-acid changes in another region of the tubulin dimer might have a similar effect on microtubule structure, much as changes that cause herbicide resistance in plants occur in both α -tubulin and β -tubulin, and not all directly affect the herbicide-binding site⁵. Plant and animal tubulins do have preferred amino acids at certain sites⁶. It is possible, then, that in the origins of climbing plant species, selective pressure resulted in the conservation of one or more amino acids in the tubulin dimer that can cause twisting. This selective pressure could have come from an intense competition for light, for example in forests, where the ability to twist and climb would have been an advantage.

Patrick J. Hussey is in the Integrative Cell Biology Laboratory, School of Biological and Biomedical Sciences, University of Durham, South Road, Durham DH1 3LE, UK.

e-mail: p.j.hussey@durham.ac.uk

1. Kerner von Marilaun, A. *The Natural History of Plants, Their Forms, Growth, Reproduction and Distribution* (transl. Oliver, F. W., with the assistance of Lady Busk & Macdonald, M. F.) (Blackie, London, 1902).
2. Thitamadee, S., Tsuchihara, K. & Hashimoto, T. *Nature* **417**, 193–196 (2002).
3. Furutani, I. *et al.* *Development* **127**, 4443–4453 (2000).
4. Adinall, S. G. & Lutkenhaus, J. *Mol. Microbiol.* **22**, 231–237 (1996).
5. Anthony, R. G. & Hussey, P. J. *Trends Plant Sci.* **4**, 112–116 (1999).
6. Burns, R. G. & Surridge, C. D. in *Microtubules* (eds Hyams, J. S. & Lloyd, C. W.) 3–31 (Wiley-Liss, New York, 1994).
7. Löwe, J., Li, H., Downing, K. H. & Nogales, E. *J. Mol. Biol.* **313**, 1045–1057 (2001).

Geology

A foot in the past

The oldest fossils of footprints ever found on land suggest that animals may have emerged from the seas much earlier than had been thought. Lobster-sized, centipede-like animals made the prints shown here about 500 million years ago. Previous fossils had indicated that animals took this step around 40 million years later.

R. B. MacNaughton and colleagues analysed about 25 rows of footprints preserved in Cambrian–Ordovician aeolian sandstone in southeastern Ontario, Canada (*Geology* **30**, 391–394; 2002). They concluded that the ripples and fine layering in the sandstone are characteristic of wind-blown sand compacted over millennia, rather than

underwater sediments, as had previously been believed.

The animals that made the tracks were about 50 cm long, had 16–22 legs, and may have dragged a tail behind them. They were probably euthycarcinoids — relatives of modern centipedes, and similar in appearance to overgrown woodlice.

Given the fossils' age, it is unlikely that the creatures lived on land. They probably ventured ashore to mate and lay eggs, to escape predators or to scavenge for food.

But if animals did forsake the seas 500 million years ago, they may have found little vegetation to eat. The only known fossils of land plants that are as old as these



footprints are remains of algal mats. So the tracks appear to contradict the prevailing hypothesis that animals colonized the land to exploit leafy resources. A single finding is insufficient evidence, however, and sandstone rocks of this age are notoriously difficult to date. More examples will be needed before palaeontologists can rewrite the textbooks.

Tom Clarke

Physiology

The pitfalls of power laws

Ewald R. Weibel

The conundrum of how an animal's metabolic rate is related to its size continues to exercise biologists. A possible solution that takes into account differences during rest and exercise deserves attention.

Why does a mouse burn six times more energy per unit body mass per minute than a human? The intuitive answer is to keep warm. In 1883, Rubner¹ found that the metabolic rate depended on the body's surface-to-volume ratio, or $M^{2/3}$, where M is the body mass, thus supporting the notion of temperature regulation. But in 1932, Kleiber² claimed that the oxygen consumption of a resting body scaled instead with $M^{3/4}$. There is no obvious reason for the exponent to have this value, however, and for the past 70 years investigators have either contested it or attempted to explain it. On page 166 of this issue, Darveau *et al.*³ present a new concept that takes this discussion a significant step forward. Whereas most previous analyses sought a single cause for what has been called the '1/4 power law', these authors argue that there is no single power-law relation between metabolic rate and body size. They find that the relation changes between rest and exercise, and they explain why.

The observation that the functional and structural properties of organisms show allometric scaling — scaling that follows some power of the body mass — has long been an intellectual challenge in biology. Some cases appear straightforward; for example, the scaling of stride frequency (the rate at which animals move their legs) with $M^{1/6}$ can be explained on the grounds of mechanical efficiency. But it is less easy to explain the allometric scaling of metabolic rate. There is still no agreement on which allometric function is correct — although the majority vote, including the textbooks, seems to favour Kleiber's 3/4 power relationship⁴. But a thorough re-examination of large data sets has concluded that the null hypothesis of a 2/3 power relation cannot be rejected⁵.

Several attempts have been made to explain the 3/4 power relation between metabolic rate and body mass. To give just one example, West, Brown and Enquist⁶ have modelled the system of blood vessels supplying oxygen to the capillaries as a fractal network and claim that this design ultimately constrains the rate of oxygen supply. Fractal structures are related to power functions, and these authors derive a compelling set of allometric exponents, including the exponent 3/4 for metabolic rate. It is indeed an attractive model, a simple unified theory as the solution for a complex system — almost too good to be true.

The trouble with the 3/4 exponent is that

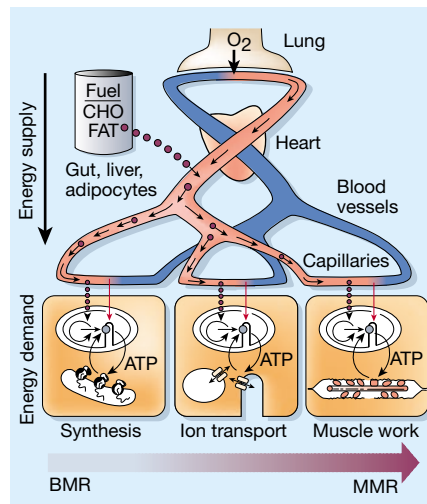


Figure 1 The allometric dependence of metabolic rate on body size has been a controversial subject, with no agreement on the value of the power-law exponent. Darveau *et al.*³ argue that the key is the shift in the body's energy demands between rest and exercise. At rest, basal metabolic rate (BMR) is determined by the energy demands of basic maintenance processes in the tissues, such as protein synthesis by ribosomes (left) and ion transport by membrane pumps (centre). When an animal exercises, muscle work through the actomyosin complex (right) places a much larger demand on the energy-supply cascade from lung to capillaries and mitochondria, and the ease with which oxygen can be transported through the cascade limits the maximal metabolic rate (MMR). The result is that BMR and MMR have different allometric dependencies on body mass, and the scaling rates can be explained only by mixing the correct 'cocktails' of energy-demanding components.

it is discussed in terms of a limitation on the oxygen supply, whereas in fact it characterizes the relation between body mass and basal metabolic rate (BMR), when the animal is using as little energy as possible and is merely sustaining basic life processes. Taylor *et al.*⁷ have found that maximal metabolic rate (MMR) induced by exercise scales with $M^{0.86}$ rather than $M^{0.75}$. A fresh approach is clearly needed, and this is provided by Darveau *et al.*³. They recognize that metabolic rate is a complex feature that results from a combination of functions, and they apply the notion of distributed control that is being used in relation to biochemical pathways⁸.

Figure 1 shows three cells performing different functions that require energy (in the form of ATP supplied by mitochondria, the cell's power-houses): protein synthesis by ribosomes; ion transport by membrane pumps; and muscle work through the actomyosin complex. Darveau *et al.*³ argue convincingly that, in going from rest to exercise, there is a shift in function from left to right. At rest, muscle activity is minimal, but during strenuous exercise some basic metabolic functions, such as digestion and protein synthesis, are run down. Thus, BMR is determined by the energy demands of the basic maintenance processes in the tissues, including, of course, the production of heat to stabilize body temperature. To supply the minimum amount of energy, blood flow is distributed to the organs according to need, much going to the brain, kidney, liver and gut, and little to the muscles⁹.

When an animal exercises, energy demand by the muscles (which account for 40% of body mass) dominates: at MMR, the demand for oxygen increases 10–15-fold, even 30–50-fold in horses. Because oxygen cannot be stored, the energy-supply cascade from lungs to capillaries and mitochondria, driven by the heart, becomes greatly stressed, with the consequence that some of the blood flow is shifted away from maintenance organs such as the kidney and gut. In contrast, fuels are stored abundantly, even in the cells; instantaneous supply is less important, eventually limiting endurance but not affecting MMR¹⁰.

Darveau *et al.*³ present a new model that takes into account all the component steps of the system, whether in parallel or in series. They calculate the allometric function for the overall metabolic rate as a weighted average and conclude that the different scaling of BMR and MMR can be explained by mixing different 'cocktails' of the components. BMR results from the combination of energy demands by different cells. However, MMR is determined by the high energy demand of muscles, and therefore the scaling is determined by the steps of the oxygen pathway⁹, through the energy-supply cascade. Interestingly, the authors estimate that BMR scales with $M^{0.73}$ and that MMR scales with $M^{0.86}$. The scaling exponent for MMR is higher because large species have greater metabolic scope (that is, a larger range of metabolic rate above BMR).

This new approach to understanding the relation between metabolic rate and body size demonstrates the importance of taking an integrated view of the functions occurring in an organism⁹. It offers a way of assessing how each of the varying functions and different design features contributes to overall performance while maintaining the integral view.

Even though we now seem to be on the right track, there are still some intriguing facts that suggest this is not the last word. For example, how do we account for the fact that



100 YEARS AGO

One of the most remarkable architectural structures in existence is the left-handed spiral staircase in the Chateau de Blois, Touraine, built during the sixteenth century from designs by Leonardo da Vinci. In a well-illustrated and thoughtful article ... Mr. Theodore Cook shows that the design of this staircase corresponds so exactly with the spirals on the common Mediterranean shell known as *Voluta vespertilio* as to leave little doubt that the artist had that shell before him as his model. The spiral on the central column of the core of the staircase corresponds exactly, for instance, with the spiral ridges on the columella of the volute, as seen in section. This of itself would be strong, although perhaps not absolutely convincing, evidence as to the origin of the design. But the staircase has also an exquisite outer balustrade, which shows a correspondence to the coils on the external spire of the shell as close as that which obtains between the interior of the staircase and the columella of the volute ... It is remarkable, however, that the spirals in the staircase run in the reverse direction to those in normal examples of the shell, that of the central shaft being left-handed instead of right-handed.

From *Nature* 8 May 1902.

50 YEARS AGO

Dr. Derek Price has recently written two articles ... in which he describes some very interesting facts brought to light by a recent examination of a manuscript entitled "The Equatorie of the Planetis" in the Perne Library, Peterhouse, Cambridge. This was written in 1392, and if, as is now thought probable, it is a hitherto unknown work of Chaucer, the manuscript is of great importance as it would provide for the first time an example of Chaucer's handwriting and also an uncorrupted specimen of his language and spelling. Various lines of evidence suggest that Chaucer was the author and not Simon Bredon, to whom it was once attributed, as it is now known that the latter died on or before 1372, the date when his will proved. Probably the strongest evidence for the authorship is found in a note adjacent to a table for the conversion of years to solar days ... The "Equatorie" appears to be a free adaptation from an Arabic or Persian source, presumably through a Latin translation.

From *Nature* 10 May 1952.

the most important energy-transfer step in metabolism (the maximum rate of oxidative ATP synthesis per unit of mitochondrial membrane) doesn't vary with body mass¹¹? What about the safety factors that seem to be built into many functions — how do they affect this kind of overall relationship? In fact, allometric relations may mask various features, such as the fact that the metabolic scope of certain athletic species of larger animals may be two to five times greater than is typical for that size of animal⁹. Thus, at MMR, the most athletic mammal, the pronghorn antelope of the Rocky Mountains, consumes oxygen at the same rate as a mouse¹² — so body size is not a universal primary determinant of energetics.

Whether Darveau *et al.*'s model³ is the ultimate wisdom remains to be seen, but it does give the field of comparative integrative physiology a new thrust. And it does away, perhaps

regrettably, with a myth — that in comparative biology a single magic number might unlock all of nature's secrets.

Ewald R. Weibel is in the Department of Anatomy, University of Berne, Bülhstrasse 26, CH-3000 Berne 9, Switzerland.

e-mail: weibel@ana.unibe.ch

1. Rubner, M. *Z. Biol.* **19**, 535–562 (1883).
2. Kleiber, M. *Hilgardia* **6**, 315–353 (1932).
3. Darveau, C. A., Suarez, R. K., Andrews, R. D. & Hochachka, P. W. *Nature* **417**, 166–170 (2002).
4. Schmidt-Nielsen, K. *Scaling: Why is Animal Size So Important?* (Cambridge Univ. Press, 1984).
5. Dodds, P. S., Rothman, D. H. & Weitz, J. S. *J. Theor. Biol.* **209**, 9–27 (2001).
6. West, G. B., Brown, J. H. & Enquist, B. J. *Science* **284**, 1677–1679 (1999).
7. Taylor, C. R. *et al. Respir. Physiol.* **44**, 25–37 (1981).
8. Srere, P. A. *Biol. Chem. Hoppe-Seyler* **374**, 833–842 (1993).
9. Weibel, E. R. *Symmorphosis* (Harvard Univ. Press, Boston, 2000).
10. Weibel, E. R. *et al. J. Exp. Biol.* **199**, 1699–1709 (1996).
11. Hoppeler, H. & Lindstedt, S. L. *J. Exp. Biol.* **115**, 355–364 (1985).
12. Lindstedt, S. L. *et al. Nature* **353**, 748–750 (1991).

Applied physics

Bridge for the terahertz gap

Carlo Sirtori

A device has been invented that produces radiation in the terahertz range. It is a considerable feat of semiconductor fabrication, and could be used in a wide range of applications.

In modern communications, information is transmitted by modulating either the amplitude or the frequency of electromagnetic waves. But not all of the electromagnetic spectral range can be exploited for this purpose, a deficiency that is addressed by Köhler *et al.* on page 156 of this issue¹.

The electromagnetic range that is used is vast, crossing the hertz span from kilo (10^3), through giga (10^9) to tera (10^{12}), and so running from radio to optical frequencies — that is, from about 10 kHz (wavelength around 30 km) to about 300 THz (wavelength around 1 μm). Depending on whether we are listening to the BBC, or making a telephone call on a mobile phone or across the ocean, we are connected through signals that are carried by radiation at 100 kHz (radiowave), 3 GHz (microwave) or 300 THz (near-infrared light). Semiconductor devices are the enabling components for all of these technologies. Oscillating circuits based on high-speed transistors efficiently produce radiation in the low-frequency region; at high frequencies, semiconductor lasers generate coherent light for telecommunication links based on optical fibres.

As shown in Fig. 1, however, the frequency ranges of the two technologies do not meet. On the one hand, transistors and other quantum devices based on electron transport (see, for instance, ref. 2) are limited to about 300 GHz (50 GHz being the rough practical limit; devices much above that are

extremely inefficient). On the other hand, the wavelength of semiconductor lasers can be extended down to only 10 μm (about 30 THz)³. Between the two technologies lies the so-called terahertz gap, where no semiconductor technology can efficiently convert electrical power into electromagnetic radiation. Köhler *et al.*¹ describe a new semiconductor laser that produces intense radiation at 4.4 THz (wavelength 67 μm), by injecting electrons into a quantum structure. There is still a long way to go before commercial devices based on this principle can be mass-produced. But this demonstration shows that they might be feasible: it could result in the bridging of the terahertz gap between transistor and laser technologies.

Like other semiconductor lasers, Köhler and colleagues' device is composed of an active region, where photons are produced, and an optical waveguide that confines the radiation. But these two fundamental constituents of a laser must meet completely different demands in the terahertz region. The main difficulties to be overcome are the high intrinsic optical losses induced by the free electrons in the material; the large size of the optical mode — imposed by the wavelength; and the short lifetime of the excited state of the laser transition, in which electrons accumulate to create the 'population inversion' necessary for lasing. Köhler *et al.* have carefully addressed each of these points.

The waveguide is an innovative hybrid

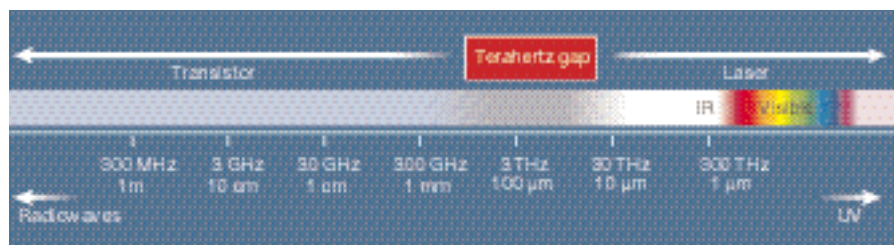


Figure 1 The terahertz gap. The gap, lying roughly between 300 GHz (0.3 THz) and 30 THz, exists because the frequencies generated by transistors and lasers, typical semiconductor devices, don't overlap. No current semiconductor technology can efficiently convert electrical power into electromagnetism in that range. But the 'heterostructure laser' produced by Köhler *et al.*¹ might, in due course, meet the demand for radiation sources at these terahertz wavelengths.

between a 'metallic microwave stripe-line' and a 'surface plasmon waveguide', a concept already used for semiconductor lasers that produce radiation in the mid-infrared range⁴. The propagation losses measured by Köhler *et al.* are of the order of 10 cm^{-1} — that is, a reduction of about a factor of three for each millimetre travelled. This is much the same as in the mid-infrared, where the wavelength is ten times smaller, and is surprisingly low: free-carrier absorption increases with the square of wavelength, so one would expect much higher optical losses at around $70\text{ }\mu\text{m}$. The trick is that the only absorbing region is a narrow layer of heavily doped semiconductor that keeps half of the radiation mode in the active region, and leaves the other half in a loss-less substrate⁵. The waveguide design principle and radiation 'mode' profile are shown in Fig. 2 of the paper on page 157.

The active region is made of many identical quantum structures connected in series. In this quantum cascade scheme⁶, electrons are selectively injected in an excited state; they then 'jump' to a lower energy state, emitting light; and finally they are recycled in the next structure in the series. To obtain laser action, most of the electrons have to be in the excited state. When, as in this case, the energy transition between the excited and lower energy states is a few meV, the electrons tend to scatter and the excited state has too short a lifetime for effective population inversion. Scattering is a function of the electron density, and Köhler *et al.* have tackled the problem by spreading the total population inversion over a sequence of 104 repeat periods in the quantum structures. In this way, the total of the electron population contributing to the optical gain is comparable to that of a conventional diode laser, but is two orders of magnitude less per period. It means, however, that the quantum structure constituting the active region is more than $10\text{ }\mu\text{m}$ thick and contains almost 1,500 semiconductor layers, each of which is only a few nanometres thick and needs precise doping. Creating such a thick and complicated structure was a considerable feat of semiconductor-crystal growth.

This is not the first semiconductor laser

that can operate in the terahertz region. More than 15 years ago, lightly doped germanium showed laser action at low (cryogenic) temperatures and under crossed electric and magnetic fields⁷. At the moment, the performance of the two types of device is comparable, but Köhler *et al.*'s heterostructure laser represents major progress because of the scope for refinement. The main difference between the two structures is in the way they are conceived. The germanium laser depends on the fundamental properties of the bulk material, which are inflexible and difficult to manipulate. The laser designed by Köhler *et al.*, by contrast, is the result of sophisticated quantum engineering, with the design of the structure determining the optical (emission frequency) and electrical properties. It is the control of device properties, and the flexibility offered by semiconductor crystal growth, which mean that improvements leading to a commercial terahertz technology may well be possible.

What applications might take advantage of such a technology? Terahertz emitters

would be in great demand for use in astronomy and studies of Earth's atmosphere — even if this is only a niche market, it is a well-defined and long-lasting one. But medical imaging, wireless telecommunications, chemical analyses using molecular spectroscopy and intra-building links (terahertz radiation can pass through construction materials) are also areas that could benefit greatly. The most enticing and profitable application is of course in telecommunications, in greatly extending the frequency band available. Nevertheless, two main points need to be kept in mind. First, the atmosphere becomes very opaque above 1 THz, absorbing radiation at these frequencies. Second, no efficient or fast terahertz detectors are yet available. The challenge is, therefore, to move to even lower frequencies — although meeting this challenge would mean moving into a competitive arena with existing transistor technology.

Finally, there is just a little *déjà vu* about the paper by Köhler *et al.*¹. It seems that the latest trends in semiconductor laser are taking us back to the origins of laser technology — back to the late 1950s, and to the microwave frequencies exploited in the invention of the maser.

Carlo Sirtori is in the Semiconductor Lasers Group, Thales Research and Technology, 91404 Orsay Cedex, France.

e-mail: carlo.sirtori@thalesgroup.com

1. Köhler, R. *et al.* *Nature* **417**, 156–159 (2002).
2. Rodwell, M. J. W. *et al.* *IEEE Trans. Electron Devices* **48**, 2606–2624 (2001).
3. Beck, M. *et al.* *Science* **295**, 301–305 (2002).
4. Sirtori, C. *et al.* *Optics Lett.* **23**, 1366–1368 (1998).
5. Ullrich, J. *et al.* *Physica B* **272**, 216–218 (1999).
6. Faist, J. *et al.* *Science* **264**, 553–556 (1994).
7. Andronov, A. A. *et al.* *Pis'ma Zh. Eksp. Teor. Fiz.* **40**, 69 (1984). *Transl. JETP Lett.* **40**, 804 (1984).

Palaeoecology

Climate records spruced up

Peter D. Moore

Pollen analysis is a valuable tool in helping to reconstruct past climatic conditions. Such studies can be informative on local as well as regional scales, as findings with fossil spruce pollen in Maine show.

Taking the past as a guide to the future is an approach used in many endeavours. One is the modelling of future climate change, and in such work, assemblages of fossil pollen are among the sources of evidence about the past. Such assemblages accumulate in the stratified sediments of lakes and peat deposits, and they are of course derived from, and reflect, the vegetation of former times.

But pollen travels far and wide, so most studies can provide only broad indications of past vegetation and lack the resolution needed to reconstruct local features of palaeoclimate. By using a series of small,

forest-enclosed hollows in New England as their source of pollen stratigraphy, however, Schaffler and Jacobson¹ have succeeded in obtaining the required spatial resolution for that region. As they report in *Journal of Ecology*, they are able to conclude that the coastal region of Maine became distinctly cooler than inland areas around 5,000 years ago, and so show the usefulness of such studies.

The value of pollen analysis as a source of information about past climatic conditions has been recognized for about a century. For most of that time, the sites most prized for such studies have been large lakes and

extensive, preferably unforested, peatlands that can reasonably be expected to have gathered their pollen from wide regions and to have been relatively little influenced by local factors. Many such sites have now been documented, so it has become possible to reconstruct the changing patterns of trees and other plants in the past, particularly during the Holocene (the past 10,000 years), and from these to deduce patterns of climate change^{2,3}.

Small sites of sedimentation and pollen preservation, say less than 100 m in diameter, especially when surrounded by forested vegetation, have long been regarded as poor sources of regional pollen because they are dominated by the surrounding local vegetation⁴. But the usefulness of such sites became apparent to archaeologists, who are often more interested in patterns of land use, and therefore in local vegetation change, than in wider issues relating to regional vegetation⁵. Their work has often exploited such small sites, or the edges of larger ones where the local impacts of land use have been more strongly felt.

Schauffler and Jacobson¹ now show that the local record of vegetation history from small depositional sites can also contribute to climatic reconstruction by adding greater spatial resolution. They have examined and analysed five sites of sedimentation all within 1 km of the coast of Maine, and two further sites from inland. Previous studies in small sites^{6,7} have shown that up to half of the pollen input is derived from trees within 50–100 m, so the record of forest history is essentially a very local one.

Records of spruce (*Picea* species) pollen proved particularly interesting in the work in Maine. This tree genus prefers cooler conditions. It was present in the region at the end of the last glacial episode but was then driven northwards as the climate became warmer in the Holocene, and as competing pine (*Pinus* species) and hemlock (*Tsuga canadensis*) forest expanded. But, from the forest-hollow pollen records, it seems that spruce survived locally in the coastal strip and expanded strongly there around 5,000 years ago, although its extension inland took place only in the past 1,000 years.

Schauffler and Jacobson interpret these local variations in terms of the persistent oceanic cooling effect near the coast, which encouraged spruce survival there. The coastal expansion 5,000 years ago can be linked with oceanographic evidence for a further cooling in the surface waters of the Gulf of Maine at that time. Inland, the influence of the maritime cooling was not sufficient to give spruce the competitive advantages it evidently gained along the coast. So it failed to survive through the Holocene but later expanded into the interior from the 'refugial' populations on the coast.

Schauffler and Jacobson's study demon-

strates the potential value of vegetation records from small, local sites in overall reconstructions of climate. It underlines the spatial complexity of vegetation changes and of individual species movements. Analyses of timberline dynamics from Chile⁸ to Sweden⁹ show that the response of vegetation to climate change is patchy in space and time, and the work in Maine indicates how important refugial populations of a species can be in permitting rapid response to changing climatic conditions. Those using fossil-pollen records in climatic modelling should certainly not neglect data from small sites of sediment accumulation. Site size does matter in palaeoecological interpretation — but

small does not mean negligible.

Peter D. Moore is in the Division of Life Sciences, King's College London, Franklin Wilkins Building, 150 Stamford Street, London SE1 9NN, UK.

e-mail: peter.moore@kcl.ac.uk

1. Schauffler, M. & Jacobson, G. L. Jr *J. Ecol.* **90**, 235–250 (2002).
2. Huntley, B. & Birks, H. J. B. *An Atlas of Past and Present Pollen Maps for Europe: 0–13,000 Years Ago* (Cambridge Univ. Press, 1983).
3. Huntley, B. & Prentice, I. C. *Science* **241**, 687–690 (1988).
4. Jacobson, G. L. Jr & Bradshaw, R. H. W. *Quat. Res.* **16**, 80–96 (1981).
5. Behre, K.-E. (ed.) *Anthropogenic Indicators in Pollen Diagrams* (Balkema, Rotterdam, 1986).
6. Jackson, S. T. & Wong, A. J. *Ecol.* **82**, 89–99 (1994).
7. Sugita, S. *J. Ecol.* **83**, 879–898 (1995).
8. Cuevas, J. G. *J. Ecol.* **90**, 52–60 (2002).
9. Kullman, L. *J. Ecol.* **90**, 68–77 (2002).

Cell biology

Keeping the genome in shape

Frank Uhlmann

Chromosomes adopt their well-known form only at a certain phase in the cell-division cycle, just before they separate. But the proteins that help shape chromosomes also seem to be at work earlier on.

Each human cell measures just micrometres in diameter but contains a staggering two metres or so of DNA, encoding all the information required to make a human being. The sheer amount of DNA poses a formidable challenge to cellular proliferation, when a cell must duplicate its genome and later (during mitosis) separate the copies into two new cells. To make things easier, the human genome comes in 46 pieces — the chromosomes — and, to ensure their safe division, these become highly compacted during mitosis by a process called chromosome condensation. One key player in this process is a protein complex known as condensin. Surprisingly, though, it seems that this complex also has other tasks. Recent work, including the paper by Aono and co-workers¹ on page 197 of this issue, shows that condensin is also important in the period between divisions — interphase — during which chromosomes are 'decondensed' and not individually visible in the nucleus.

The condensin complex is one of the most abundant protein components of mitotic chromosomes, but is largely absent from chromosomes in interphase^{2,3}. This discovery hinted that condensin might be important in mitotic chromosome condensation (Fig. 1a, overleaf), and the suspicion was confirmed in experiments using extracts from frog eggs, where chromosome condensation could be analysed *in vitro*². When condensin was removed from such extracts, chromosomes failed to condense in mitosis and instead remained in a diffuse, interphase-like state. Other experiments identified mutations in components of the condensin complex of fission yeast⁴; these

mutations likewise resulted in the failure of chromosomes to condense in mitosis, and led to severe mitotic defects. This was a striking demonstration of the disastrous consequences for genome segregation if condensation fails.

Since then, condensin has been found and characterized in all eukaryotic organisms looked at — from fission and budding yeasts to fruitflies, worms, frogs and humans. The condensin complex consists of five proteins, two of which belong to the SMC (or 'structural maintenance of chromosomes') family. Members of this family are also found in the other two domains of life, the Archaea and Bacteria; in Bacteria they are involved in genome compaction and segregation⁵. So condensin might have truly universal chromosome-shaping properties.

It now appears to have another, totally unexpected role, as Aono *et al.*¹ discovered by studying a new condensin mutation in fission yeast. Cells with this mutation were highly sensitive to DNA damage and in particular to a compound, hydroxyurea, that blocks the progression of DNA replication. Could this be an indirect consequence of the general genomic instability suffered by cells that have mutations in condensin? The authors find that this is probably not the case; rather, condensin has a direct role in monitoring DNA replication.

The authors reveal that although condensin concentrates on fission-yeast chromosomes in mitosis, smaller amounts of the complex remain associated with chromosomes in interphase, when DNA is replicated. If DNA replication is blocked, cells normally recognize the problem and

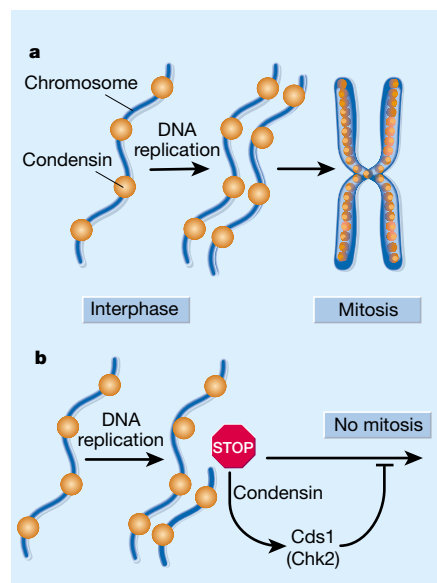


Figure 1 Dual role of the condensin protein complex during the cell-division cycle. **a**, Condensin is a major component of chromosomes during mitosis (nuclear division; right), and is required for chromosomes to become more compact relative to the looser form they adopt in interphase (the period between nuclear divisions). However, Aono *et al.*¹ now show that some condensin is also present on interphase DNA (left). **b**, These authors find that, in interphase, condensin is important if chromosome replication is blocked. Normally, cells are then prevented from entering mitosis by a cellular response called the replication checkpoint (involving the Cds1/Chk2 protein), as shown here. But fission yeast with a mutation in condensin fail to activate this pathway and begin a fatal mitosis, even if replication is incomplete¹. Condensin must therefore be important for monitoring chromosome replication.

initiate a response pathway, the 'replication checkpoint' (Fig. 1b). This pathway will try to get replication going again, and will also prevent cells from dividing until replication is completed⁶. One component of this pathway is an enzyme called Cds1 (also known as Chk2), which is activated when replication stalls. But Aono *et al.*¹ find that their mutant yeast fail to trigger this response. Cds1 is not activated, and the cells enter mitosis even though DNA replication is incomplete — with fatal consequences. So condensin is apparently needed on interphase chromosomes to sense replication blocks and to switch on the replication checkpoint. To make things a little more complicated, the mutant condensin not only results in a failure to respond to replication blocks, but also seems to generate such blocks itself¹.

How does the condensin that is bound to interphase DNA work? Does its role in preventing and sensing replication blocks rely on its ability to help chromosomes to condense, albeit to a lesser extent than in mitosis? Or does it have a second, independent mode of action? Some remarkable hints about how condensin functions have come from *in vitro* studies⁷, which showed that this complex can introduce twists in DNA that increase its tendency to fold back on itself. More recently, low-resolution images of the condensin complex have been obtained^{8,9}. But the significance of these observations for mitotic chromosome condensation is unclear. And how condensin might help cells to prevent and recognize replication blocks is still more mysterious.

A question that might be easier to answer is whether condensin acts throughout the genome in interphase, or is recruited to specific sites at which replication is blocked or damage is encountered. So far, only a few DNA sites to which condensin binds during interphase have been characterized. One of them is the rDNA in budding yeast¹⁰. Such

DNA is a repetitive region of the genome that encodes the RNA of ribosomes — cellular protein-synthesizing machines. rDNA also contains sequences that stop DNA replication from moving against the direction in which genes are transcribed. Whether the binding of condensin to such natural replication barriers is related to its role in triggering the replication checkpoint remains to be seen.

Finally, yet another function of condensin during interphase has become clear recently. In fruitflies, condensin binds to chromosomal elements that regulate the transcription of neighbouring genes, and is required to keep these genes 'silent'¹¹. Condensin mutants in budding yeast similarly fail to maintain the silent state of a gene that helps to dictate mating type¹². Do these events reflect the diverse functions of a single complex, or different effects of a single function? Either way, it will be fascinating to work out the molecular mechanisms by which the condensin complex acts to safeguard and shape our precious blueprint — DNA. ■

Frank Uhlmann is at the Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3PX, UK.

e-mail: frank.uhlmann@cancer.org.uk

1. Aono, N., Sutani, T., Tomonaga, T., Mochida, S. & Yanagida, M. *Nature* **417**, 197–202 (2002).
2. Hirano, T. & Mitchison, T. J. *Cell* **79**, 449–458 (1994).
3. Saitoh, N., Goldberg, I. G., Wood, E. R. & Earnshaw, W. C. *J. Cell Biol.* **127**, 303–318 (1994).
4. Saka, Y. *et al.* *EMBO J.* **13**, 4938–4952 (1994).
5. Britton, R. A., Chi-Hong Lin, D. & Grossman, A. D. *Genes Dev.* **12**, 1254–1259 (1998).
6. Boddy, M. N. & Russell, P. *Curr. Biol.* **11**, R953–R956 (2001).
7. Kimura, K., Rybenkov, V. V., Crisona, N. J., Hirano, T. & Cozzarelli, N. R. *Cell* **98**, 239–248 (1999).
8. Yoshimura, S. H. *et al.* *Curr. Biol.* **12**, 508–513 (2002).
9. Anderson, D. E., Losada, A., Erickson, H. P. & Hirano, T. *J. Cell Biol.* **156**, 419–424 (2002).
10. Freeman, L., Aragon-Alcaide, L. & Strunnikov, A. *J. Cell Biol.* **149**, 811–824 (2000).
11. Lupo, R., Breiling, A., Bianchi, M. E. & Orlando, V. *Mol. Cell* **7**, 127–136 (2001).
12. Bhalla, N., Biggins, S. & Murray, A. W. *Mol. Biol. Cell* **13**, 632–645 (2002).

Daedalus

Brighter clouds

The evidence for global warming is patchy but alarming. By the time we know whether carbon dioxide is the villain it may be too late. Clouds are the natural reflectors of sunlight. Even a small improvement in their reflectivity would save us. An improvement might be made if the top layer of a cloud consisted, not of droplets, but of hollow bubbles about the same size. Each would then have four reflecting surfaces rather than two, and would reflect sunlight much better. A bubble, of course, is lighter than a droplet: so a cloud with both droplets and bubbles should have its bubbles concentrated at the top, just where they are needed. And once a bubble has formed, it won't seal into a solid droplet again. Desorbed air will get into the internal gas-space. Furthermore, says Daedalus, a trace of surface-active solute may form a stabilizing internal monolayer.

One of the sources of CO₂ is the flaring of natural gas that cannot be got to market. Daedalus wants to blow this waste into a detergent solution. With well-designed bubblebers, this would produce a mass of bubbles each about the size of a cloud droplet. Methane is lighter than air, so these tiny bubbles would rise. With luck, they would join the world's clouds. A natural thermocline restricts the height of clouds, and should control their bubbly top layer too. Fortunately, the initial methane is only 'catalytic'. The bubbles will soon start to lose it by outward emission, at the same sort of rate as they gain air by absorption and water vapour by internal release. Water vapour, of course, is almost as light as methane. So in due course all the clouds in the atmosphere will be topped, not by water droplets, but by tiny bubbles.

This new cloud cover should reflect sunlight more efficiently than the old one. Even a minor programme of methanation will get the process going, and once methane bubbling has started in earnest, we should soon see the bubbly-top cloud layer growing. The speed and efficiency of the scheme will be confirmed by reflectance measurement from satellites.

Daedalus cannot guess the useful life of hollow cloud droplets. Like normal cloud drops, they may evaporate completely. Or they may accrete together and fall as a fizzy sort of rain. But with luck they should be stable enough, and last long enough, to reduce or reverse the menace of global warming. Luckily, little detergent need be introduced into the hydrological cycle. Last time, non-biodegradable detergents gave problems.

David Jones

Eukaryotic diversity in Spain's River of Fire

This ancient and hostile ecosystem hosts a surprising variety of microbial organisms.

The Rio Tinto, known by the Phoenicians as 'Ur-yero', or 'River of Fire', because of its deep red colour and high acidity, flows through the world's largest pyritic belt in southwestern Spain. Surprisingly, eukaryotic microbes are the principal contributors of biomass in this hostile river, which has a pH of 2 and contains much higher concentrations of heavy metals than are typically found in fresh waters. Here we show that the Rio Tinto shows an unexpected degree of eukaryotic diversity and includes new lineages that we have identified by sequence analysis of genes encoding small-subunit ribosomal RNAs. The diversity of these eukaryotes is much greater than that of prokaryotes, whose metabolism is responsible for the extreme environment.

The Rio Tinto (Fig. 1) is 90 km long. Unlike other extreme acidic environments, the low pH and high concentrations of heavy metals in this vast ecosystem (3–20 g iron per litre of water, for example) predate 5,000 years of human mining activity, with iron-oxidizing bacteria having existed in the river basin some 300,000 years ago¹.

Morphology-based investigations first alerted us to the possible variety of eukaryotic lineages in the Rio Tinto. Our molecular analysis reveals the extent of this eukaryotic diversity — which is quite unexpected in such an acidic and highly metal-rich environment — as well as several taxa and cosmopolitan organisms that have never previously been reported in extreme habitats (Fig. 2).

Eukaryotic algae contribute at least 60%

of the biomass in the Rio Tinto², forming conspicuous patches on the river's surface. Our molecular analysis detected many further, non-photosynthetic lineages, such as ciliates, cercomonads, vahlkampfiid amoebae, stramenopiles and fungi, which have all previously escaped detection³.

The phylogenetic placement of several sequences was consistent with the idea that they represent new eukaryotic lineages. Two groups of clones branched as possible sister taxa to amoeboid lineages^{4,5}: the first of these (clones RT5iin14 and RT5iin16) branched at the base of the animal–fungal–nuclearioid radiation; the second (RT5iin21 and RT5iin44) branched with the filose amoeba (*Filamoeba nolandii*), with moderately strong bootstrap support (83%). However, the long-branched nature of these sequences makes it difficult to position them precisely in a large phylogeny. The clone RT5iin25 may represent a new stramenopile lineage, with only weak bootstrap values supporting its relationship to *Cafeteria roenbergensis*.

The evolutionary distances between acidophilic species that occur in the river and their neutrophilic relatives offer insight into adaptation to an extremely acidic environment. We detected clones from the Rio Tinto that are closely related to neutrophilic species such as *Chlamydomonas noctigama*, *Chlorella minutissima* and *Colpidium campylum*. Adaptations associated with the transition from a neutral to an acidic environment must occur relatively rapidly when measured on evolutionary timescales.



Figure 1 The Rio Tinto at La Palma del Condado, Spain. The river's red colour is due to the high concentration of iron dissolved in its extremely acidic waters.

We know from microscopic observations that rotifers also inhabit the river, as do heliozoa and other types of amoeba. The wide eukaryotic diversity associated with low pH and high concentrations of heavy metals indicates that adaptation to such adverse conditions must be more easily accomplished than was previously thought. As eukaryotes rarely flourish in extreme environments, our results challenge former ideas about the phylogenetic range of organisms that are capable of living in extreme environments.

Linda A. Amaral Zettler*, **Felipe Gómez†**, **Erik Zettler‡§**, **Brendan G. Keenan||**, **Ricardo Amils†§**, **Mitchell L. Sogin***

*Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA
e-mail: sogin@evl5.mbl.edu

†Centro de Astrobiología, INTA-CSIC, Torrejón de Ardoz 28850, Spain

‡Sea Education Association, PO Box 6, Woods Hole, Massachusetts 02543, USA

§Centro de Biología Molecular, UAM-CSIC, Cantoblanco, Madrid 28049, Spain

||Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut 06269, USA

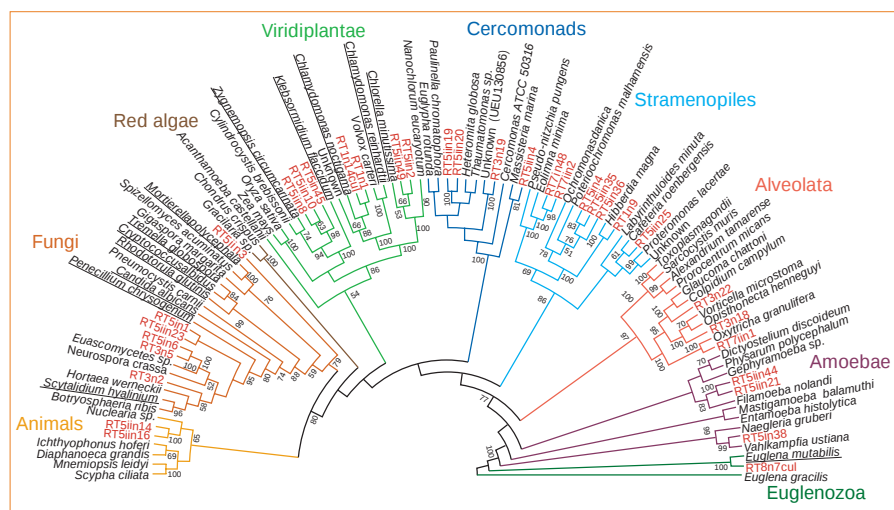


Figure 2 Phylogeny based on minimum-evolution analyses of ribosomal RNAs using a likelihood model. Bold letters indicate environmental clones. RT, sequences from Rio Tinto species; cul, sequences from cultured species. Underlined taxa represent genera that have been identified in the river by microscopy. Sampling sites: RT1, La Palma; RT3, Berrocal Upper; RT5i, the Origin, black filamentous biofilm; RT5ii, the Origin, green filamentous biofilm; RT7i, Anabel's Garden, green biofilm; RT7ii, Anabel's Garden, yellow biofilm. Bootstrap values are shown next to branches. Sequences have been deposited at GenBank under accession numbers AY082969–AY083001.

- González-Toril, E. et al. in *Biohydrometallurgy: Fundamentals, Technology and Sustainable Development* (eds Ciminelli, V. S. T. & García, O.) 639–650 (Elsevier, Amsterdam, 2001).
- Lopez-Archilla, A. I., Marin, I. & Amils, R. *Geomicrobiol. J.* **11**, 223–233 (1994).
- Durán, C., Marin, I. & Amils, R. in *Biohydrometallurgy and the Environment: Towards the Mining of the 21st Century* (eds Amils, R. & Ballester, A.) 521–530 (Elsevier, Amsterdam, 1999).
- Amaral Zettler, L. A., Nerad, T. A., O'Kelly, C. J. & Sogin, M. L. *J. Eukaryot. Microbiol.* **48**, 293–297 (2001).
- Amaral Zettler, L. A. et al. *Protist* **151**, 275–282 (2000).

Competing financial interests: declared none.

Brain damage

Neglect disrupts the mental number line

A popular metaphor for the representation of numbers in the brain is the 'mental number line', in which numbers are represented in a continuous, quantity-based analogical format^{1,2}. Here we show that patients with hemispatial neglect³ misplace the midpoint of a numerical interval when asked to bisect it (for example, stating that five is halfway between two and six), with an error pattern that closely resembles the bisection of physical lines⁴. This new form of representational neglect constitutes strong evidence that the mental number line is more than simply a metaphor, and that its spatial nature renders it functionally isomorphic to physical lines.

Although few of us think of numbers in spatial terms⁵, psychophysical evidence suggests that the mental representation of numbers has a spatial nature, possibly with a left–right orientation⁶. First, both humans⁷ and primates⁸ are quicker and more accurate in selecting the greater of two numbers when the numerical distance between them is large than when it is small. Second, smaller numbers are responded to faster with the left hand and larger ones with the right hand^{6,9}.

Patients with unilateral neglect resulting from a (right) parietal lesion show a spatial deficit for left-side stimuli, which also

extends to mental images³. When asked to mark the midpoint of a line, they miss the midpoint and place it to the right. The misplacement increases as a function of line length, with a crossover effect (leftward displacement) for very short lines⁴. We therefore investigated whether these patients show the same form of distortion when bisecting the mental number line.

We tested four right-brain-damaged patients with persistent left neglect, four right-brain-damaged patients without spatial neglect, and four healthy control subjects. All patients showed intact numerical and arithmetic skills. In particular, they achieved near-perfect scores in subtraction and number comparison, two tasks that tap the core deficit in parietal acalculia¹⁰ after lesion of the language-dominant hemisphere. Participants heard two numbers that defined a number interval and were asked to state the midpoint number without making calculations. The size of the interval could be three (for example, 1–3), five, seven or nine. Each interval was presented using the units (for example, 1–5), the teens (11–15) and the first tens (21–25), giving a total of 48 trials. Two neglect patients and all control subjects repeated the task using reverse presentation of the same number pairs (for example, 9–1).

Neglect patients made many errors (individuals' error rates: O.V., 39.58%; G.Z.A., 37.5%; G.Z.E., 31.25%; G.S., 35.41%). Performance was significantly affected by the size of the number interval (Fig. 1a). For the three patients who made

errors at the smallest interval, their midpoint answers were significantly shifted to the 'left' in the number line (for example: interval, 11–13; suggested midpoint, 10). For longer intervals, right-shifted errors were more common (for example, 11–19; 17); this phenomenon increased as a function of interval size. Reverse presentation produced the same pattern, showing that the number line is canonically orientated¹¹ in a left-to-right manner. Control subjects made few errors and their performance was not affected by interval size (Fig. 1b). The magnitude of numbers used did not influence the performance of any of the groups of participants.

The systematic midpoint shift observed here is indeed produced by neglect — one control patient had a comparable error rate (31%) to that of the neglect patients, but his errors were not modulated by interval size and were randomly distributed between leftward and rightward shifts (16 left, 14 right). Moreover, the pattern of errors shown by neglect patients does not resemble that of parietal acalculics^{10,12} (the only other patients known to experience difficulty with this task), who fail in a variety of tasks that require manipulation of numerical quantities and can perform extremely poorly in number bisection (error rates can be as high as 77%; ref. 12).

The performance of neglect patients closely mirrors their difficulty in bisecting physical lines⁴. This demonstrates the spatial nature of the mental number line and its striking functional isomorphism to physical lines. Although most people focus on symbolic aspects of numbers, and few seem to be aware of the intimate relationship between numbers and visuo-spatial representations, thinking of numbers in spatial terms (as has been reported by great mathematicians¹³) may be more efficient because it is grounded in the actual neural representation of numbers.

Marco Zorzi *†, Konstantinos Priftis*, Carlo Umiltà*

**Dipartimento di Psicologia Generale, Università di Padova, via Venezia 8, 35131 Padova, Italy*
e-mail: marco.zorzi@unipd.it

†Università Vita-Salute San Raffaele, via Olgettina 58, 20132 Milano, Italy

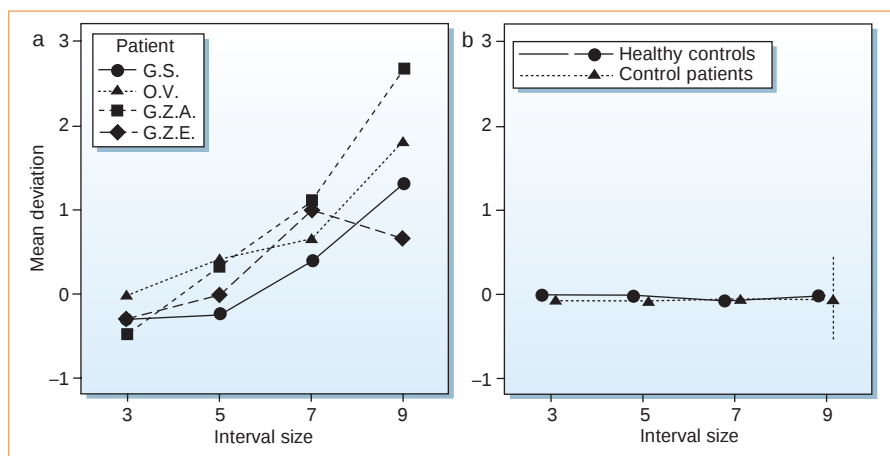


Figure 1 Displacement error with respect to the midpoint (0) of a numerical interval as a function of interval size. Positive values indicate rightward shifts of the midpoint along the number line; negative values represent leftward shifts. **a**, Performance of neglect patients (mean age, 63 yr; education, 9 yr). All had single, hypodense brain lesions limited to the right hemisphere (fronto-temporo-parietal lesion, $n=3$; fronto-parietal lesion, $n=1$; mean time since lesion, 78 days). Regression analyses revealed that the performance of all patients was significantly affected by interval size (G.S., $\beta=0.40$, $P<0.001$; O.V., $\beta=0.54$, $P<0.001$; G.Z.A., $\beta=0.62$, $P<0.001$; G.Z.E., $\beta=0.36$, $P<0.01$). For G.S., G.Z.A. and G.Z.E., the occurrence of a leftward midpoint shift for the smallest interval was significant (16 out of 18 trials, $P=0.001$, binomial test). The magnitude of the numbers used was not a significant factor in the regressions (G.S., $\beta=0.09$, n.s.; O.V., $\beta=0.08$, n.s.; G.Z.A., $\beta=0.08$, n.s.; G.Z.E., $\beta=0.04$, n.s.). **b**, Mean performance ($\pm$ standard error) of control patients (mean age, 58 yr; education, 5.2 yr) and healthy controls (mean age, 63 yr; education, 9 yr). Two control patients and three healthy controls made no errors (0 out of 96 trials). One healthy control and one control patient made some errors (error rates, 5.2% and 12.5%, respectively); these were all shifted to the left of the midpoint. The performance of the control patient who made a considerable number of errors (error rate, 31.2%) was not spatially biased (16 leftward shifts, 14 rightward) and was not affected by interval size ($\beta=0.07$, n.s.) or numerical magnitude ($\beta=0.01$, n.s.).

- Restle, F. J. *Exp. Psychol.* **83**, 274–278 (1970).
- Dehaene, S., Dupoux, E. & Mehler, J. *J. Exp. Psychol. Hum. Percept. Perform.* **16**, 626–641 (1990).
- Bisiach, E. & Vallar, G. in *Handbook of Neuropsychology* (eds Boller, F. & Grafman, G.) 459–502 (Elsevier, Amsterdam, 2000).
- Marshall, J. & Halligan, P. *Cortex* **25**, 503–515 (1989).
- Seron, X., Pesenti, M., Noël, M. P., Deloche, G. & Cornet, J. A. *Cognition* **44**, 159–196 (1992).
- Dehaene, S., Bossini, S. & Giraux, P. *J. Exp. Psychol. Gen.* **122**, 371–396 (1993).
- Moyer, R. & Landauer, T. *Nature* **215**, 1519–1520 (1967).
- Brannon, E. & Terrace, H. *Science* **282**, 746–749 (1998).
- Bächtold, D., Baumüller, M. & Brugger, P. *Neuropsychologia* **36**, 731–735 (1998).
- Dehaene, S., Dehaene-Lambertz, G. & Cohen, L. *Trends Neurosci.* **21**, 355–361 (1998).

11. Caramazza, A. & Hillis, A. E. *Nature* **346**, 267–269 (1990).
12. Dehaene, S. & Cohen, L. *Cortex* **33**, 219–250 (1997).
13. Hadamard, J. *The Mathematician's Mind: The Psychology of Invention in the Mathematical Field* (Princeton Univ. Press, New Jersey, 1996).

Competing financial interests: declared none.

Electrochemistry

Building on bubbles in metal electrodeposition

In the electrodeposition of metals, a widely used industrial technique, bubbles of gas generated near the cathode can adversely affect the quality of the metal coating. Here we use phase-contrast radiology with synchrotron radiation to witness directly and in real time the accumulation of zinc on hydrogen bubbles. This process explains the origin

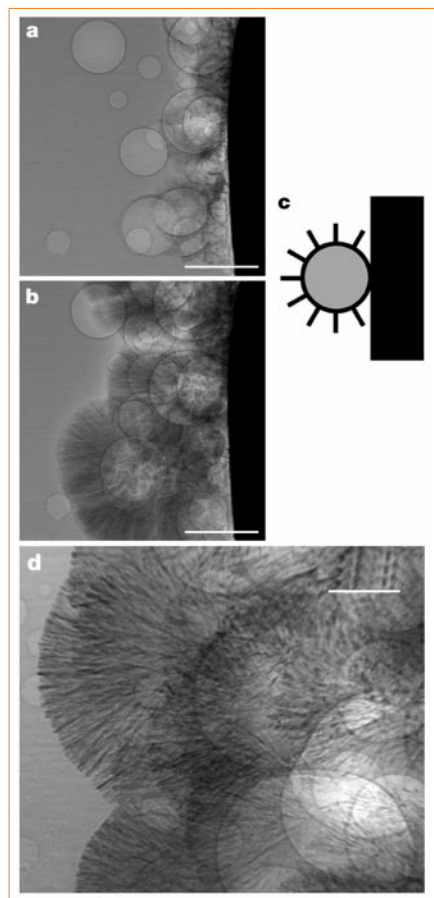


Figure 1 Phase-contrast microradiographs showing the growth of zinc on hydrogen bubbles. **a, b**, Images (taken 6 s apart) showing growth of zinc dendrites. **c**, Diagram of radial dendritic growth along the electric-field lines. **d**, Image showing the microstructure of the dendrites. Radiographs were taken at the Pohang Light Source in Korea and at the Synchrotron Radiation Research Center in Hsinchu, Taiwan. Deposition was carried out in a mini electrochemical cell with a small (about 5 mm) thickness of electrolyte solution between the windows to reduce X-ray absorption, and a variable (about 8 mm) electrode gap. The electrode was covered with epoxy, which prevented conduction, except for the X-ray-detection window. The base solution was 4.8 M KCl and 2.2 M ZnCl₂. Scale bars, 300 μm (**a, b**) and 200 μm (**d**).

of the bubble-shaped defects that are common in electrodeposited coatings.

Gas bubbles are an important problem in electrodeposition coating. For example, too much hydrogen typically gives rise to porous films with poor adhesion^{1,2}. Surface treatments are often used to solve this problem^{3–5}.

The bubble phenomenon has eluded experimental clarification for decades because of a lack of probes with adequate temporal and spatial resolution and sufficient penetration to show the pattern and development of the coating's microstructural detail.

We have used a new real-time technique known as phase-contrast radiology^{6,7} to demonstrate that zinc can grow directly on gas bubbles, leading to the formation of voids in the coating. The technique, which generates images that reflect differences in refractive index, is more effective than conventional absorption-based radiology.

Phase-contrast radiology depends on coherent X-rays from synchrotron sources^{6,7}. Even with the best sources, however, real-time studies with high spatial resolution are difficult or impossible to carry out. We have solved this problem by using unmonochromatized synchrotron X-rays (that is, X-rays with limited longitudinal coherence)^{6,7}.

Previously, research on bubbles in electrodeposition processes has relied mainly on quantitative measurements of hydrogen formation and *ex situ* characterization of coatings. The presence of bowl- and tube-shaped voids suggested negative bubble fingerprints⁸, but no direct evidence corroborated this possibility.

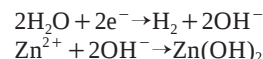
Instead, our results directly link the coating morphology with the hydrogen bubbles. For example, Fig. 1a, b shows the simultaneous formation of bubbles and zinc dendrites. The images reveal that the metal coating is formed directly on the bubble surface, with the zinc dendrites growing out radially (Fig. 1c). Figure 1d shows the microstructure of these dendrites.

Under varying experimental conditions, we found zinc growing on bubbles not only as dendrites but also as films and in other forms. We did not observe any deposition on bubbles that were not connected to the electrode (Fig. 1a), a finding that rules out electrodeless deposition.

On the basis of these results, we developed a model to explain the effect of bubbles on metal-coating quality. In this model, a bubble is generated and adheres to the electrode surface; zinc nucleation then occurs at high-surface-energy sites on the electrode–bubble interface, producing microstructures that perturb the local electric field. This enables zinc deposition to extend laterally over the bubble surface, covering the entire bubble. Dendrites then grow radially, following the configuration of the electric field (Fig. 1c).

The metallic film and the dendrite structure are able to stabilize the bubble-related shape. When the deposited metal becomes sufficiently dense, permanent voids are left inside the film, giving rise to defects in the coating.

This mechanism requires a charge-transfer medium. Zinc hydroxide (which, like most metal hydroxides, is an electrical conductor⁹) is the most likely candidate. Zn(OH)₂ can be generated as follows:



Reduction of hydrogen increases the concentration of zinc hydroxide near the cathode surface^{10–13}. A similar process probably occurs near the bubble surface, resulting in deposition of zinc. We tested this idea by injecting hydrogen into the solution to form hydrogen–electrolyte–electrode interfaces. No zinc was deposited on these interfaces, ruling out the unlikely possibility that the bubbles act directly as charge carriers.

Phase-contrast radiology can be put to several other uses — for example, we have generated dynamic maps of density in the electrolytic solution. Our discovery that metal coatings can literally be built on bubbles finally confirms what has long been suspected.

W. L. Tsai*, P. C. Hsu*, Y. Hwu*, C. H. Chen†, L. W. Chang†, J. H. Je‡, H. M. Lin§, A. Groso||, G. Margaritondo||

*Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan

†China Steel Corporation, 1 Chung-Kang Road, Kaohsiung 81233, Taiwan

‡Department of Materials Sciences, Pohang University of Science and Technology, Pohang 790-784, Republic of Korea

§Department of Materials Engineering, Tatung University, Taipei 10452, Taiwan

||École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

e-mail: giorgio.margaritondo@epfl.ch

1. Coleman, D. H., Popov, B. N. & White, R. E. *J. Appl. Electrochem.* **28**, 889–894 (1998).
2. Monev, M. et al. *J. Appl. Electrochem.* **28**, 1107–1112 (1998).
3. Manso, M., Jiménez, C., Mrant, C., Herrero, P. & Martínez-Duart, J. M. *Biomaterials* **21**, 1755–1761 (2000).
4. Chen, J. M. & Wu, J. K. *Plating Surface Finish* **10**, 74–77 (1992).
5. Zamanzadeh, M., Allam, A., Kato, C., Ateya, B. & Pickering, H. W. *J. Electrochem. Soc.* **129**, 285–289 (1982).
6. Hwu, Y. et al. *J. Appl. Phys.* **86**, 4613–4618 (1999).
7. Margaritondo, G. & Tromba, G. *J. Appl. Phys.* **85**, 3406–3408 (1999).
8. Yumoto, H., Kinase, Y., Ishihara, M., Baba, N. & Kamei, K. *J. Surface Sci. Soc. Jpn* **17**, 43–48 (1996).
9. Jayashree, R. S. & Kamath, P. V. *J. Power Sources* **93**, 273–278 (2001).
10. Brenner, A. *Electrodeposition of Alloys: Principles and Practices* vols 1, 2 (Academic, New York, 1963).
11. Elkhatibi, F., Sarret, M. & Muller, C. *J. Electroanal. Chem.* **404**, 45–53 (1996).
12. Younan, M. *J. Appl. Electrochem.* **30**, 55–60 (2000).
13. Velichenko, A. B., Portillo, J., Alcobé, X., Sarret, M. & Muller, C. *Electrochimica Acta* **46**, 407–414 (2000).

Competing financial interests: declared none.

Chemokine-receptor genes and AIDS risk

Schliekelman *et al.*¹ have provided a model to quantify the speed at which HIV-resistance haplotypes can become enriched in a susceptible population through a delay in the onset of AIDS, permitting greater lifetime reproduction and the selection of AIDS-delaying haplotypes. But we question their conclusion¹ that there could be a rapid evolution of resistance to AIDS onset in some African populations if the current HIV epidemic persists, as this depends on an untested assumption: that variant forms of the chemokine-receptor-5 (CCR5) gene impart selective advantages or disadvantages in Africa that are comparable to those reported for African Americans^{2–6}. Here we test this premise in a large Ugandan population, and find that CCR5 variants are not associated with HIV/AIDS disease risk in Africa — the origin and centre of the current AIDS pandemic. This gene may therefore not be subject to rapid evolutionary change as a result of the HIV epidemic in Africa.

We genotyped 1,149 Ugandan HIV-1-positive individuals with previously measured CD4 counts⁷ at the CCR5 locus. No CCR5 promoter allele was associated either with infection in a case-control study or with disease progression to death in a Cox regression analysis (Tables 1, 2). Neither the guanine–thymine substitution at position –2554, which characterizes the HHD and HHC haplotypes associated with maximal disease progression in African Americans², nor the guanine–adenine change at position –2459, which is consistently associated with accelerated progression in Caucasians^{3,6,8,9}, was significantly associated with the rate of disease progression to death in Uganda.

We derived haplogroups and haplotypes from the different CCR5-promoter genotypes²; analysis of 440 individuals with ascertainable haplogroups showed no overall correlation with the rate of disease

Table 1 Allele frequencies of CCR5 and CCR2 variants in HIV-positive and uninfected individuals

	CCR2 64I	–2733G	–2554T	–2459G	–2135C	–2132T	–2086G	–1835T
Allele frequency (HIV-positive)	0.203	0.145	0.163	0.468	0.424	0.12	0.047	0.346
Allele frequency (controls)	0.214	0.124	0.183	0.464	0.456	0.121	0.052	0.317
P value (allelic comparison, d.f. = 1)	0.59	0.3	0.26	0.87	0.26	0.94	0.67	0.21
P value (genotypic comparison, d.f. = 2)	0.25	0.5	0.07	0.85	0.26	0.73	0.88	0.19
P value (Cox regression, d.f. = 2)	0.34	0.28	0.28	0.365	0.94	0.70	0.60	0.18

Allele frequencies and P values for promoter single-nucleotide polymorphisms² in CCR5 in HIV-positive cases and uninfected controls. P values denote the significance of results. Cox modelling of a possible effect on the rate of disease progression to death included the genotype, ethnic group (Bantu or non-Bantu) and CD4 count at entry into the study.

progression. In particular, the haplogroup HHA/HHF*2, which is most clearly associated with disease retardation in African Americans², was the most prevalent haplogroup in Ugandans but was not associated with altered disease progression.

Furthermore, the valine–isoleucine amino-acid change at position 64 of CCR2, which is in linkage disequilibrium with the CCR5 promoter and has been associated with delayed disease progression^{4,10}, was not associated with disease retardation in Ugandans (see www.well.ox.ac.uk/hill/ramaley). We also examined nine CCR5 coding-region variants^{11–13}, and found variation at only five sites, three of which have a variant allele frequency of <0.1%. None of these sites was associated with either HIV infection or rate of disease progression (see www.well.ox.ac.uk/hill/ramaley).

Our examination of over 1,100 HIV-positive and over 300 HIV-negative Ugandans represents the largest study of the genetics of HIV-1 susceptibility in an African population to date. The alleles and haplotypes identified in North America and Europe as conferring resistance or susceptibility to HIV are not associated with altered susceptibility to infection or altered rate of disease progression in this African population.

There are several factors that differ between these regions and which may account for these findings. In Uganda, the predominant HIV-1 clades are A and D — as opposed to B in Europe and North America — and transmission is predominantly heterosexual, unlike in almost all published

studies of the genetics of HIV/AIDS susceptibility. Furthermore, the types of opportunistic and chronic infection found in HIV-positive individuals in Africa differ significantly from those in Europe and North America. Africans also show more genetic diversity than Caucasians, and these two groups show differences in genotype frequencies at loci that may interact epistatically with CCR5 variants.

Whatever the mechanisms that underlie this interpopulation heterogeneity, our findings call for caution in extrapolating associations between immunogenetic factors and infectious disease from one population to another. They also indicate that inferences regarding the influence of natural selection by HIV-1 on CCR5 genetic diversity in Africa¹ need revision.

Patricia A. Ramaley*, Neil French†, Pontiano Kaleebu‡, Charles Gilks†, James Whitworth‡, Adrian V. S. Hill*

*Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

e-mail: adrian.hill@well.ox.ac.uk

†Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK

‡Uganda Virus Research Institute/Medical Research Council (UK) Programme on HIV/AIDS, PO Box 49, Entebbe, Uganda

Table 2 Cox-regression analysis of CCR5 haplogroups and disease progression

	CD4 count	P value	β	Exp(β)	95% confidence interval for exp(β)	
					Lower	Upper
HHA/HHE	49	0.172	–0.556	0.574	0.259	1.273
HHA/HHF2	49	0.619	–0.186	0.83	0.398	1.729
HHE/HHF2	28	0.621	–0.205	0.815	0.361	1.837
HHF2/HHG1	26	0.111	–0.685	0.504	0.217	1.17
HHF2/HHF2	25	0.109	–0.722	0.486	0.201	1.174
HHA/HHH	24	0.341	–0.403	0.668	0.291	1.533
HHA/HHA	23	0.402	–0.361	0.697	0.3	1.621
HHE/HHF1	23	0.648	0.194	1.214	0.529	2.785
HHA/HHF1	20	0.728	0.151	1.163	0.498	2.717
HHA/HHG1	20	0.443	–0.402	0.669	0.24	1.866

Cox-regression results include ethnic group (Bantu or non-Bantu), CD4 count at entry into the study, and the 10 most common CCR5-promoter haplogroup genotypes². P > 0.05 in all cases and the 95% confidence intervals for exp(β) span 1.0.

- Schliekelman, P., Garner, C. & Slatkin, M. *Nature* **411**, 545–546 (2001).
- Gonzalez, E. *et al. Proc. Natl Acad. Sci. USA* **96**, 12004–12009 (1999).
- Martin, M. P. *et al. Science* **282**, 1907–1911 (1998).
- Mummid, S. *et al. Nature Med.* **4**, 786–793 (1998).
- Kostrikis, L. G. *et al. J. Virol.* **73**, 10264–10271 (1999).
- An, P. *et al. AIDS* **14**, 2117–2122 (2000).
- French, N. *et al. Lancet* **355**, 2106–2111 (2000).
- McDermott, D. H. *et al. Lancet* **352**, 866–870 (1998).
- Clegg, A. O. *et al. AIDS* **14**, 103–108 (2000).
- Smith, M. *et al. Science* **277**, 959–965 (1997).
- Ansari-Lari, M. A., Liu, X. M., Metzker, M. L., Rut, A. R. & Gibbs, R. A. *Nature Genet.* **16**, 221–222 (1997).
- Carrington, M. *et al. Am. J. Hum. Genet.* **61**, 1261–1267 (1997).
- Petersen, D. C. *et al. AIDS* **15**, 171–177 (2001).

brief communications is intended to provide a forum both for brief, topical reports of general scientific interest and for technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on *Nature's* website (www.nature.com) or on request from nature@nature.com

Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2)

S. D. Bentley*, K. F. Chater†, A.-M. Cerdeño-Tárraga*, G. L. Challis†‡, N. R. Thomson*, K. D. James*, D. E. Harris*, M. A. Quail*, H. Kieser†, D. Harper*, A. Bateman*, S. Brown*, G. Chandra†, C. W. Chen§, M. Collins*, A. Cronin*, A. Fraser*, A. Goble*, J. Hidalgo*, T. Hornsby*, S. Howarth*, C.-H. Huang§, T. Kieser†, L. Larke*, L. Murphy*, K. Oliver*, S. O'Neil*, E. Rabinowitsch*, M.-A. Rajandream*, K. Rutherford*, S. Rutter*, K. Seeger*, D. Saunders*, S. Sharp*, R. Squares*, S. Squares*, K. Taylor*, T. Warren*, A. Wietzorrek†, J. Woodward*, B. G. Barrell*, J. Parkhill* & D. A. Hopwood†

* The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

† John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK

‡ Department of Chemistry, University of Warwick, Coventry CV4 7AL, UK

§ Institute of Genetics, National Yang-Ming University, Shih-Pai, Taipei 112, Taiwan

Streptomyces coelicolor is a representative of the group of soil-dwelling, filamentous bacteria responsible for producing most natural antibiotics used in human and veterinary medicine. Here we report the 8,667,507 base pair linear chromosome of this organism, containing the largest number of genes so far discovered in a bacterium. The 7,825 predicted genes include more than 20 clusters coding for known or predicted secondary metabolites. The genome contains an unprecedented proportion of regulatory genes, predominantly those likely to be involved in responses to external stimuli and stresses, and many duplicated gene sets that may represent 'tissue-specific' isoforms operating in different phases of colonial development, a unique situation for a bacterium. An ancient synteny was revealed between the central 'core' of the chromosome and the whole chromosome of pathogens *Mycobacterium tuberculosis* and *Corynebacterium diphtheriae*. The genome sequence will greatly increase our understanding of microbial life in the soil as well as aiding the generation of new drug candidates by genetic engineering.

Nutritionally, physically and biologically, soil is a particularly complex and variable environment. Streptomycetes are among the most numerous and ubiquitous soil bacteria¹. They are crucial in this environment because of their broad range of metabolic processes and biotransformations. These include degradation of the insoluble remains of other organisms, such as lignocellulose and chitin (among the world's most abundant biopolymers), making streptomycetes central organisms in carbon recycling. Unusually for bacteria, streptomycetes exhibit complex multicellular development, with differentiation of the organism into distinct 'tissues': a branching, filamentous vegetative growth gives rise to aerial hyphae bearing long chains of reproductive spores. The importance of streptomycetes to medicine results from their production of over two-thirds of naturally derived antibiotics in current use (and many other pharmaceuticals such as anti-tumour agents and immunosuppressants), by means of complex 'secondary metabolic' pathways. Furthermore, streptomycetes are members of the same taxonomic order (Actinomycetales) as the causative agents of tuberculosis and leprosy (*Mycobacterium tuberculosis* and *M. leprae*), the genomes of which have been sequenced^{2,3}. Much should be learned about these pathogens from genome-level comparisons with harmless saprophytic relatives such as streptomycetes.

Streptomyces coelicolor A3(2) is genetically the best known representative of the genus⁴. The single chromosome is linear with a centrally located origin of replication (*oriC*) and terminal inverted repeats (TIRs) carrying covalently bound protein molecules on the free 5' ends. Replication proceeds bidirectionally from *oriC*, leaving a terminal single-stranded gap on the discontinuous strand after removal of the last RNA primer. An unusual process of 'end-patching' by DNA synthesis primed from the terminal protein fills the gap⁵. Studies of many streptomycetes, including most notably a close relative of the A3(2) strain, *Streptomyces lividans* 66, established further novelties. More than a million base pairs (bp) of DNA

at either end of the chromosomes can undergo extensive deletions and amplifications without compromising viability under laboratory conditions⁶, and early comparisons of linkage maps established that most streptomycetes show conservation of gene order (synteny) in the core region⁷. Here, we report the use of an ordered cosmid library⁸ to sequence the *S. coelicolor* genome. The strain used, M145, is a prototrophic derivative of strain A3(2) lacking its two plasmids (SCP1, linear, 365 kb, AL590463, AL590464; and SCP2, circular, 31 kb, AL645771, which have been sequenced separately).

Genome structure

General features of the chromosome sequence are shown in Table 1 and Fig. 1. At 8,667,507 bp it is the largest completely sequenced bacterial genome. The *oriC* and *dnaA* gene are about 61 kb left of the centre, at 4,269,853–4,272,747 bp. Like many other microbial genomes, there is a slight bias (55.5%) towards coding sequences on the leading strand. Although less pronounced than for most other eubacterial chromosomes, there is a discernible decrease in the GC bias around *oriC*, thought to be related to DNA replication⁹. In contrast to all other bacterial genomes studied to date, however, the

Table 1 General features of the chromosome

Component of chromosome	Property
Total size	8,667,507 bp
Terminal inverted repeat	21,653 bp
G + C content	72.12%
Coding sequences	7,825
...of which pseudogenes	55
Coding density	88.9%
Average gene length	991 bp
Ribosomal RNAs	6 × (16S–23S–5S)
Transfer RNAs	63
Other stable RNAs	3

S. coelicolor chromosome displays a downward rather than an upward shift, indicating a small bias towards C on the leading strand.

Coding density is largely uniform across the chromosome, with only a slight decrease in the distal regions. The distribution of different types of genes reveals, however, a central core comprising approximately half the chromosome and a pair of chromosome arms (Fig. 1). Nearly all genes likely to be unconditionally essential—such as those for cell division, DNA replication, transcription, translation and amino-acid biosynthesis—are located in the core (exceptions tend to be duplicate genes). In contrast, ‘contingency’

loci coding for probable non-essential functions, such as secondary metabolites, hydrolytic exoenzymes, the conservons (conserved operons) and ‘gas vesicle’ proteins (see below), lie in the arms. Curiously, this biphasic structure of the chromosome does not align with the position of *oriC*. The core appears to extend from around 1.5 Mb to 6.4 Mb, giving uneven arm lengths of approximately 1.5 Mb (left arm) and 2.3 Mb (right arm). The difference in arm lengths may reflect some gross rearrangement or different rates of DNA accumulation in each arm. The fact that *oriC* is roughly central suggests some selective pressure for such positioning.

Streptomyces coelicolor and *M. tuberculosis* are both actinomycetes

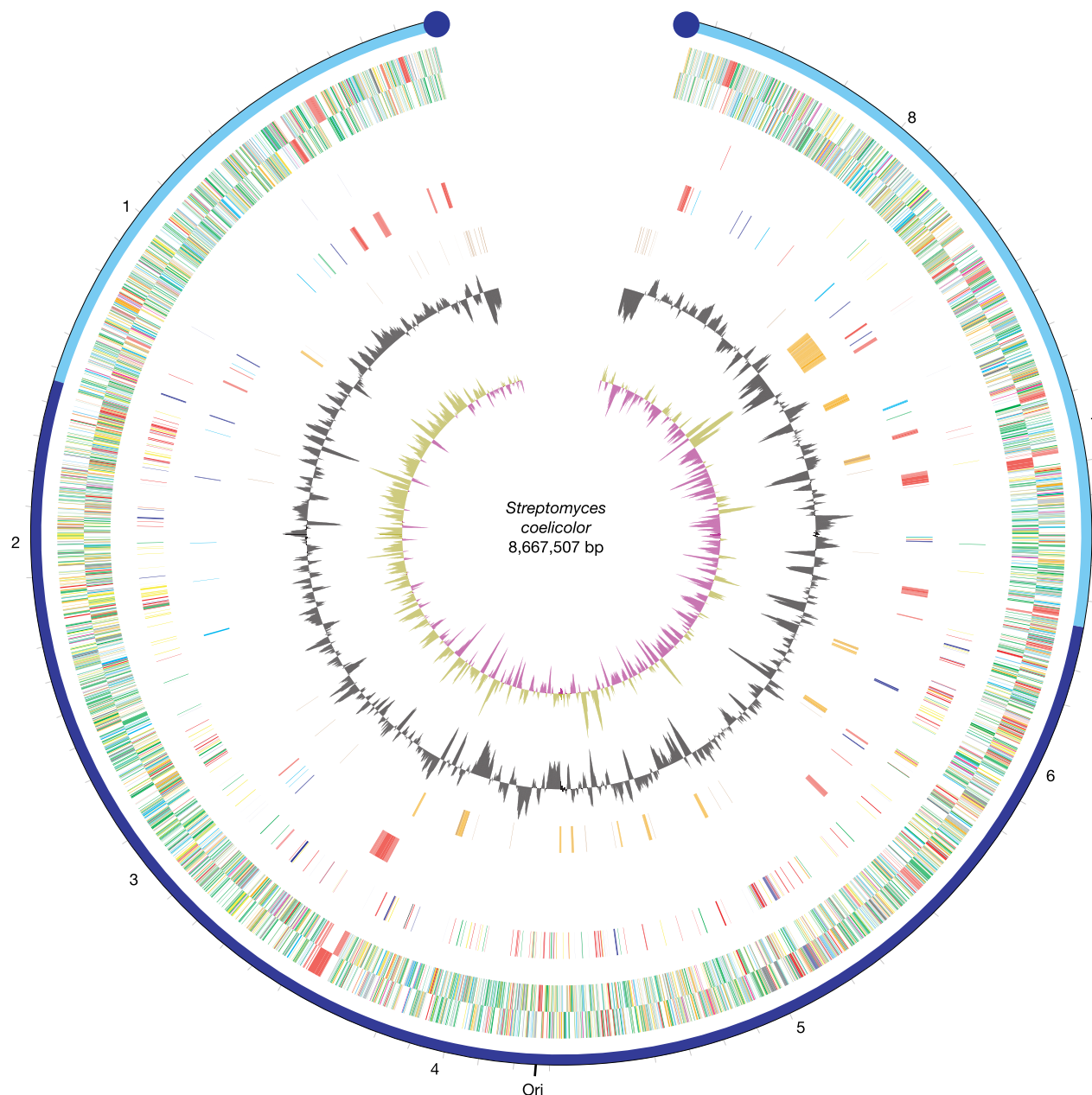


Figure 1 Circular representation of the *Streptomyces coelicolor* chromosome. The outer scale is numbered anticlockwise (to correspond with the previously published map⁸) in megabases and indicates the core (dark blue) and arm (light blue) regions of the chromosome. Circles 1 and 2 (from the outside in), all genes (reverse and forward strand, respectively) colour-coded by function (black, energy metabolism; red, information transfer and secondary metabolism; dark green, surface associated; cyan, degradation of large molecules; magenta, degradation of small molecules; yellow, central or intermediary metabolism; pale blue, regulators; orange, conserved hypothetical; brown, pseudogenes;

pale green, unknown; grey, miscellaneous); circle 3, selected ‘essential’ genes (for cell division, DNA replication, transcription, translation and amino-acid biosynthesis, colour coding as for circles 1 and 2); circle 4, selected ‘contingency’ genes (red, secondary metabolism; pale blue, exoenzymes; dark blue, conservon; green, gas vesicle proteins); circle 5, mobile elements (brown, transposases; orange, putative laterally acquired genes); circle 6, G + C content; circle 7, GC bias ((G – C/G + C), khaki indicates values >1, purple <1). The origin of replication (Ori) and terminal protein (blue circles) are also indicated.

but have very different lifestyles. Their genomes reveal much similarity at the level of individual gene sequences, and many similar gene clusters. Global comparison showed perceptible higher-order synteny as well, shown as a dot plot in Fig. 2a. A prominent feature is the central broken diagonal cross pattern formed by the regions of synteny. This broken-X pattern is commonly seen in comparisons of related bacteria and the breaks are attributed to multiple inversions centred on *oriC*¹⁰. Normally, synteny extends over the whole of the compared chromosomes; however, for the comparison between *S. coelicolor* and *M. tuberculosis*, the broken-X pattern correlates only with the core of the *S. coelicolor* chromosome. Therefore this region and the whole *M. tuberculosis* chromosome must have had a common ancestor, with the chromosome arms of *S. coelicolor* consisting of subsequently acquired DNA. The syntenic regions mainly comprise genes concerned with primary cellular functions. The most strongly conserved is the gene cluster coding for the subunits of respiratory chain NADH dehydrogenase (systematic gene numbers SCO4562–4575). Functions/proteins coded for by other regions of synteny include the origin of replication (SCO3873–3892), urease activity (SCO1231–1236), pyrimidine biosynthesis (SCO1472–1488), arginine biosynthesis (SCO1570–1580), pentose phosphate pathway/tricarboxylic acid cycle (SCO1921–1953), histidine and tryptophan biosynthesis (SCO2034–2054), cell division (SCO2077–2092) and ribosomal proteins (SCO4701–4724).

The genome of the pathogenic actinomycete *Corynebacterium diphtheriae* has been sequenced recently (<http://www.sanger.ac.uk/>

Projects/C_diphtheriae/). Comparison with the *S. coelicolor* chromosome gives a similar pattern to that for *M. tuberculosis*, with the regions of synteny covering the entire *C. diphtheriae* chromosome and just the *S. coelicolor* core region (Fig. 2b). The syntenic regions again correspond to genes coding for primary cellular functions and several of these regions are common to all three chromosomes. *Mycobacterium tuberculosis* and *C. diphtheriae* have more extensive synteny than either has with *S. coelicolor* (Fig. 2c), reflecting taxonomic groupings: *C. diphtheriae* and *M. tuberculosis* are in the suborder Corynebacterineae of the actinomycetes, whereas *S. coelicolor* is in the Streptomycineae.

By investigating regions of unusual DNA content and/or genes with sequence similarity to those from known mobile genetic elements, we designated 14 regions as potentially recently laterally acquired insertions (See Supplementary Information). By far the largest insertion contains 148 genes and is located at a transfer RNA gene¹¹; as well as many hypothetical genes, it includes genes for heavy metal resistance (SCO6835–6837) and secondary metabolite production (SCO6827). Six other inserted regions have plasmid function genes in common with the integrative plasmid pSAM2 of *Streptomyces ambofaciens*¹². Four of these pSAM2-like integrants appear to have inserted within a tRNA gene, including two that are adjacent to secondary metabolic clusters (calcium-dependent antibiotic (CDA), SCO3250–3262; *whiE*, SCO5327–5350). Notably, 11 of the 14 insertions lie to the right of *oriC*, correlating with the greater variation in DNA composition in the right half of the chromosome (Fig. 1).

Putative transposase genes are found throughout the chromosome in intact, truncated and frame-shifted forms. Many are associated with the multi-gene integrations described above. For the remainder, there is a particular concentration at the sub-TIR regions, 35–95 kb from the ends (Fig. 1). This indicates a tolerance to insertion events in these regions and thus offers a possible route for chromosome expansion. Of the 78 predicted transposase coding sequences, five are within transposons (one of which codes for a possible antibiotic resistance protein (SCO0107)), 31 form simple insertion elements and the remainder are not bounded by inverted repeats. Most fall into five families, suggesting a degree of intra-chromosomal transposition. Such events offer a route for gene duplication. Two of the insertion elements mark the inner boundaries of the TIRs, suggesting a possible role in their maintenance.

A plethora of proteins

With 7,825 predicted genes, the *S. coelicolor* chromosome has an enormous coding potential. This figure compares with 4,289 genes in the Gram-negative bacterium *Escherichia coli*; 4,099 in the Gram-positive spore-former *Bacillus subtilis*; 6,203 in the lower eukaryote *Saccharomyces cerevisiae*; and a predicted 31,780 in humans (<http://www.ebi.ac.uk/genomes/>). The genome contains almost twice as many genes as that of *M. tuberculosis*. This large number of genes reflects both a multiplicity of new protein families and an expansion within known families when compared with other bacteria (further information is available at http://www.sanger.ac.uk/Projects/S_coelicolor/). Many protein families that are significantly expanded in *S. coelicolor* are involved in regulation, transport and degradation of extracellular nutrients (Table 2).

The genome shows a strong emphasis on regulation, with 965 proteins (12.3%) predicted to have regulatory function. Discovery of so many regulators extends the observation that the proportion of regulatory genes increases with bacterial genome size¹³. There is a clear preference for certain regulator groups. Sigma factors act by binding to and affecting the promoter specificity of the RNA polymerase core enzyme, thus directing the selective transcription of gene sets. *Streptomyces coelicolor* codes for a remarkable 65 sigma factors (the next highest number so far found is 23 in *Mesorhizobium loti*, with a genome size of 7.6 Mb¹⁴), of which 45 are 'ECF' (extra-cytoplasmic function) sigma factors (41 from family 13

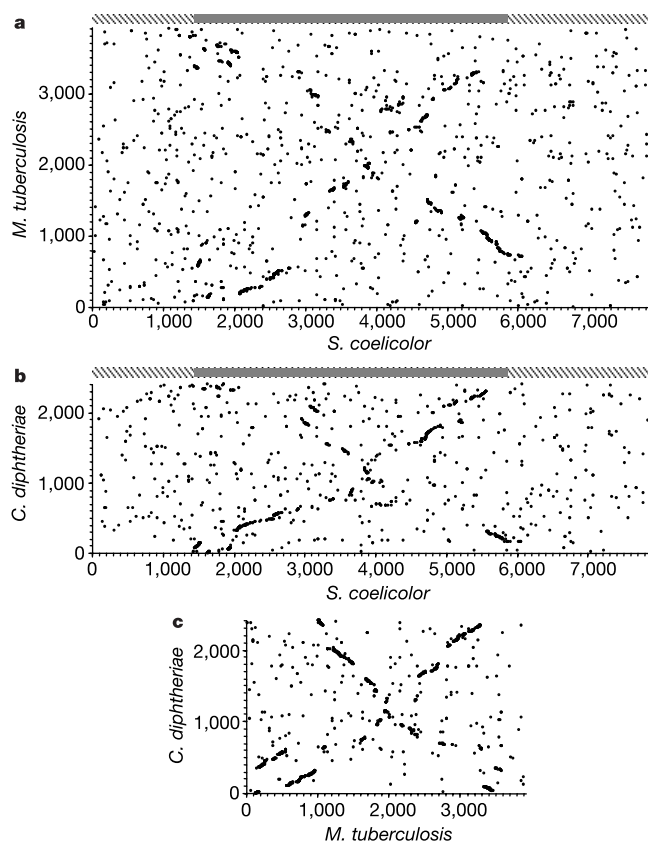


Figure 2 Comparison of chromosome structure for *S. coelicolor* versus *M. tuberculosis* (a), *S. coelicolor* versus *C. diphtheriae* (b) and *M. tuberculosis* versus *C. diphtheriae* (c). Axes represent the proteins coded for in the order in which they occur on the chromosomes. For each genome, DnaA is centrally located. Dots represent a reciprocal best match (by FASTA comparison⁵⁰) between protein sets. The bars above plots a and b indicate the core (solid, SCO1440–5869) and arm (hatched) regions of the *S. coelicolor* chromosome.

alone; Table 2). Previously described ECF sigma factors (in *S. coelicolor*) respond to external stimuli and activate genes involved in disulphide stress, cell-wall homeostasis and aerial mycelium development¹⁵. Most of the other sigma factors fall into a single group (family 54, Table 2). Within this is a sub-group peculiar to Gram-positive bacteria, most of which have a single member; however, *B. subtilis* has three, controlling forespore development and the general stress response, and *S. coelicolor* has at least eight, many of them involved in responses to various stresses¹⁶. The numerous potentially stress-responsive sigma factors may account for the independent regulation of diverse stress response regulons in *S. coelicolor*¹⁷. Although widely distributed among bacteria, the atypical, enhancer-dependent sigma-54 and its cognate activators¹⁸ are absent.

Streptomyces coelicolor also has abundant two-component regulatory systems where typically, in response to an extracellular stimulus, an integral membrane sensor protein phosphorylates a response regulator, causing it to bind to specific promoter regions and thus activate or repress transcription. We identified 85 sensor kinases and 79 response regulators, including 53 sensor-regulator pairs. The genome also codes for many members of previously described regulator families such as LysR, LacI, ROK, GntR, TetR, IclR, AraC, AsnC and MerR. The TetR family regulators in *S. coelicolor* form several subfamilies, often containing few or no members from the other genomes analysed. Furthermore, there is a group (family 86, Table 2) of 25 putative DNA-binding proteins that has no members from outside *S. coelicolor* and may constitute a new *Streptomyces*-specific family of regulators. Also notable is the presence of 44 putative serine/threonine protein kinases (family 6.1, Table 2). Examples of these typically eukaryotic regulators are now known to occur in many bacteria, but in much smaller numbers.

Reflecting its many interactions with the complex soil environment, *S. coelicolor* has 614 proteins (7.8%) with predicted transport function. A large proportion of these are of the ABC transporter type, including 81 typical ABC permeases and 141 ATP-binding proteins (24 of which are fused to membrane-spanning domains).

Transporters for which the substrate is predictable include those for sugars, amino acids, peptides, metals and other ions. There are also several possible drug efflux proteins. Import of specific substrates would in part be facilitated by the 75 putative surface-anchored substrate-binding proteins of *S. coelicolor*.

The ability of *S. coelicolor* to exploit nutrients in the soil is abundantly demonstrated by our prediction of 819 potentially secreted proteins (10.5%). Secreted hydrolases are particularly numerous (for example, family 7 (Table 2), which is over-represented in *S. coelicolor*). They include 60 proteases/peptidases, 13 chitinases/chitosanases, eight cellulases/endoglucanases, three amylases and two pectate lyases. As well as the complete Sec protein translocation system, *S. coelicolor* seems to contain the machinery and cognate signal sequences for the recently discovered TAT (twin arginine transport) system for exporting pre-folded proteins¹⁹ (T. Palmer, personal communication).

A marked example of multiple paralogues in *S. coelicolor* is a four-gene cluster that we named the conservon (for conserved operon). In the 13 such clusters (*cvnA*, *B*, *C*, *D*, *1-13*) there is unidirectional transcription and often overlap of translational start and stop codons, suggesting an operon structure. The only other known *cvn* cluster is present in *M. tuberculosis*. The protein products form distinct and exclusive families (Table 2; families 178, 177, 214, 180; *CvnA*, *B*, *C* and *D*, respectively). The first gene codes for a probable membrane protein weakly resembling sensor kinases, the fourth codes for a possible ATP/GTP-binding protein, and the other two are of unknown function. In four of the clusters the immediate downstream gene codes for a predicted cytochrome P-450.

Paralogous enzymes may sometimes represent isozymes active at different stages in the developmental cycle. One such example is the differential activities of duplicate gene clusters for glycogen synthesis in the vegetative and aerial mycelium²⁰. Here we highlight a further five examples of paralogues for metabolic enzymes in *S. coelicolor*. (1) Two gene clusters code for enzymes of the pentose phosphate pathway (SCO1935–1939 and SCO6657–6663). (2) Four loci for tryptophan biosynthesis (SCO2036–2043, SCO2117,

Table 2 Occurrence of a selection of protein families in six related genomes

Majority description*	Sco	Mtu	Cdi	Bsu	Mlo	Eco
ECF sigma factor (13)	41 (0.52)	10 (0.25)	7 (0.29)	7 (0.17)	16 (0.23)	1 (0.02)
Sigma factor (54)	14 (0.17)	3 (0.07)	2 (0.08)	8 (0.19)	3 (0.04)	4 (0.09)
Two-component sensor kinase (1.3)	27 (0.34)	8 (0.20)	5 (0.20)	12 (0.29)	41 (0.60)	20 (0.46)
Two-component sensor kinase (1.15)	6 (0.07)	0	0	0	0	0
Two-component regulator (1.6)	50 (0.63)	2 (0.05)	5 (0.20)	9 (0.21)	5 (0.07)	7 (0.16)
Two-component regulator (1.5)	24 (0.30)	11 (0.28)	6 (0.24)	13 (0.31)	21 (0.31)	14 (0.32)
Serine/threonine protein kinase (6.1)	44 (0.56)	13 (0.33)	5 (0.20)	8 (0.19)	14 (0.20)	8 (0.18)
Regulator (LacI) (2.4)	33 (0.42)	1 (0.02)	2 (0.08)	12 (0.29)	15 (0.22)	13 (0.30)
Regulator (ROK) (36)	23 (0.29)	3 (0.07)	3 (0.12)	3 (0.07)	6 (0.08)	7 (0.16)
Regulator (TetR) (112)	18 (0.22)	1 (0.02)	0	0	0	0
Regulator (KorSA/GntR) (2.9)	10 (0.12)	0	0	0	0	0
Regulator (WhiB-like)	8 (0.10)	4 (0.10)	3 (0.12)	0	0	0
DNA-binding (86)	25 (0.31)	0	0	0	0	0
ABC transport (ATP-binding) (2.1.3)	27 (0.34)	4 (0.10)	8 (0.33)	6 (0.14)	3 (0.04)	3 (0.06)
Transport (permease) (2.3)	36 (0.45)	4 (0.10)	1 (0.04)	8 (0.19)	25 (0.37)	3 (0.06)
Transport (sugar) (2.2)	36 (0.45)	4 (0.10)	1 (0.04)	8 (0.19)	26 (0.38)	3 (0.06)
Transport (substrate-binding) (1.8)	35 (0.44)	4 (0.10)	1 (0.04)	6 (0.14)	24 (0.35)	3 (0.06)
Integral membrane (59)	14 (0.17)	14 (0.35)	3 (0.12)	2 (0.04)	0	0
Membrane ATPase (42)	13 (0.16)	12 (0.30)	6 (0.24)	4 (0.09)	3 (0.04)	1 (0.02)
Secreted hydrolase (7)	100 (1.27)	19 (0.48)	8 (0.33)	21 (0.51)	8 (0.11)	9 (0.20)
Secreted hydrolase (7.3)	17 (0.21)	0	0	0	1 (0.01)	0
Secreted chitinase (7.8)	5 (0.06)	0	0	0	0	0
Secreted protease (191)	10 (0.12)	3 (0.07)	0	0	0	0
Secreted protease (7.6)	8 (0.10)	0	0	0	0	0
Secreted cellulase (7.4)	7 (0.08)	1 (0.02)	0	1 (0.02)	0	0
Secreted hypothetical (17)	25 (0.31)	11 (0.28)	8 (0.33)	13 (0.31)	8 (0.11)	9 (0.20)
Hypothetical (63)	25 (0.31)	0	0	6 (0.14)	0	0
Conservon (Cvn1–4; 178, 177, 214, 180)	13 (0.16)	1 (0.02)	0	0	0	0
Hypothetical (204)	13 (0.16)	0	0	0	0	0
Hypothetical (19)	12 (0.15)	0	0	0	0	0

Numbers indicate absolute number of proteins from each genome in each family and the percentage of the total proteins in that genome in parentheses. Family number is indicated in parentheses in the majority description column. The hierarchical numbering system reflects use of higher BlastP thresholds to break large complex families into discrete subfamilies. Complete data are available from http://www.sanger.ac.uk/Projects/S_coelicolor/. Sco, *S. coelicolor*; Mtu, *M. tuberculosis*; Cdi, *C. diphtheriae*; Bsu, *B. subtilis*; Mlo, *M. loti*; Eco, *E. coli*.

*Groupings are based on sequence similarity, so individual families do not necessarily include all representatives of each type of protein in each genome (see Methods).

SCO2147, SCO3211–3214) include two *trpC*, two *trpD* and three *trpE* genes. A *trpCDGE* locus is within the gene cluster for production of CDA²¹, a peptide antibiotic that contains tryptophan (in the 'unnatural' D form). The local cluster may ensure adequate tryptophan for CDA biosynthesis at the appropriate stage in the life cycle, independently of the needs of protein synthesis. (3) Five homologues of *fabH* code for a dedicated ketosynthase for the first step in fatty acid biosynthesis (condensation of acetyl-coenzyme A (CoA) with malonyl-CoA to yield acetoacetyl-CoA). One of the five (SCO2388) is in the main fatty acid biosynthetic operon and is essential²². Three of the other four *fabH* homologues (SCO5888, 3246, 1271) are in gene clusters for secondary metabolism: the *red* and *cda* clusters, and a cluster of unknown product (see below). At least the first two clusters determine molecules with fatty acid components, and the presence of *fabH* paralogues makes it highly probable that some of the steps in their biosynthesis use dedicated enzymes, rather than sharing enzymes functioning in primary metabolism²³. (4) Three clusters code for a typical four-subunit respiratory nitrate reductase (SCO0216–0219, SCO4947–4950, SCO6532–6535), indicating the importance of a capacity for micro-aerobic growth in what was classically regarded as an obligate aerobe. (5) Flexibility in respiration is further indicated by a second (partial) copy of the operon coding for subunits of the respiratory chain NADH dehydrogenase (SCO4599–4608).

Unexpectedly, there are two gene clusters (SCO0649–0658, SCO6499–6508) similar in sequence and gene order to an operon from *Halobacterium* sp. that is involved in the production of gas vesicle proteins, including the eight genes essential for this pheno-

type²⁴. Many overtly water-living bacteria use gas vesicles as flotation devices, but the only previous occurrence of gas vesicle genes (but not so far of the vesicles themselves) in a soil organism is in *Bacillus megaterium*²⁵. The benefit of gas vesicles to *Streptomyces* is unknown, but perhaps such buoyancy devices would allow spores to remain at the oxygen-rich surface during dispersal and germination in waterlogged soil.

Many genes for secondary metabolism

Chromosomal gene clusters specifying the biosynthesis of the aromatic polyketide antibiotic actinorhodin, the so-called RED complex of red oligopyrrole prodiginine antibiotics, and the non-ribosomal peptide CDA had previously been analysed^{12,27}, as had the *whiE* cluster of genes coding for a type II polyketide synthase for a grey spore pigment²⁸. The genome sequence reveals a further 18 clusters that would code for enzymes characteristic of secondary metabolism (Fig. 3). These include type I modular and both type I and type II iterative polyketide synthases (PKSs), chalcone synthases, non-ribosomal peptide synthetases (NRPSs), terpene cyclases, and others. The distribution of the clusters on the chromosome seems non-random, with some preponderance in the arms, but more especially in a region near the right-hand core-arm boundary (Fig. 1). Comparison with similar clusters from other organisms and the application of recently developed sequence analysis tools have, in some cases, provided insight into the probable structure of the end products determined by these genes. For example, using predictive models for substrate amino-acid recognition^{29,30}, the two NRPSs coded for by SCO0492 and SCO7681–

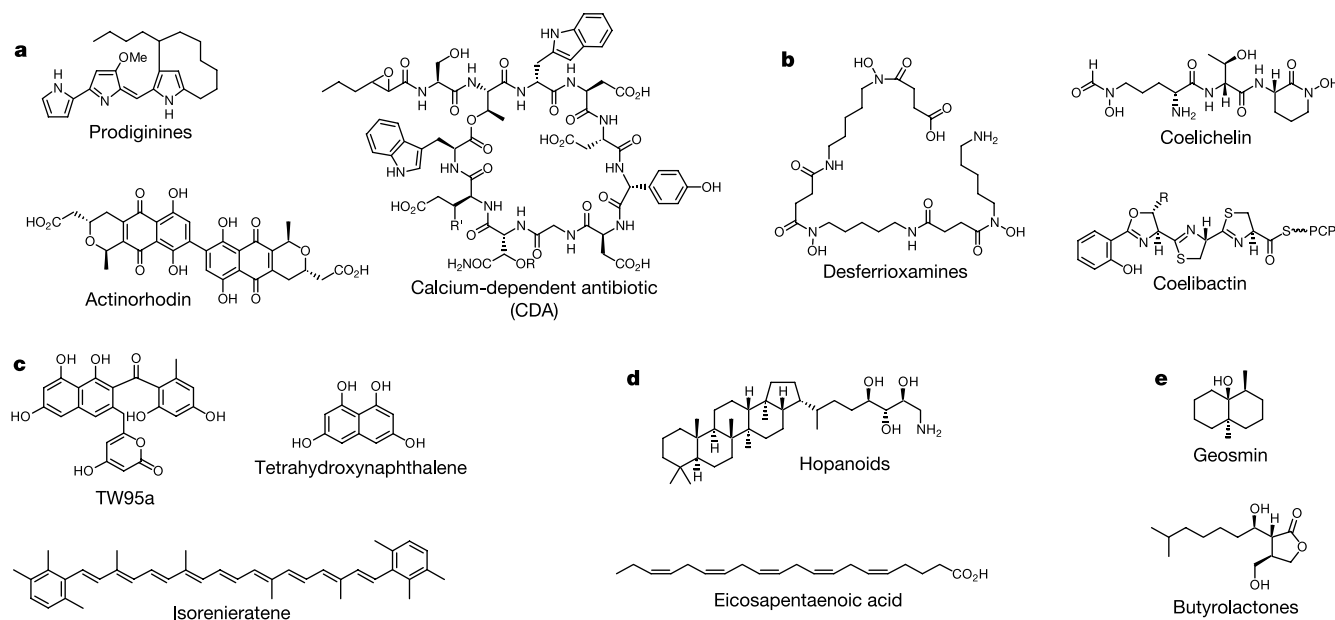


Figure 3 Secondary metabolites known or predicted to be made by *S. coelicolor* A3(2), grouped according to their putative function. These are: antibiotics (**a**), siderophores (**b**), pigments (**c**), lipids (**d**) and other molecules (**e**). The chromosomal locations of the gene clusters are: actinorhodin, SCO5071–5092; prodiginines (mixture of butyl-*meta*-cycloheptylprodiginine (shown) and undecylprodiginine), SCO5877–5898; CDA complex (CDA1, R = OP₃H₂, R' = H; CDA2, R = OP₃H₂, R' = Me; CDA3b, R = OH, R' = H; CDA4b, R = OH, R' = Me), SCO3210–3249; desferrioxamines (mixture of desferrioxamine G₁ (shown) and desferrioxamine E), SCO2782–2785; coelichelin, SCO0489–0499; coelibactin (structure is that predicted for a late intermediate attached to the PCP domain in the last module of the coelibactin NRPS; R = H/Me, the complete structure cannot be predicted as the regioselectivity of several methyltransferases, a cytochrome P-450 and an oxidoreductase coded for by genes in the cluster cannot be deduced), SCO7681–7691; TW95a (structure is the product obtained from heterologous expression of the *whiE* minimal PKS and the *whiE-ORFIV* genes; the structure of the grey

spore pigment has not been elucidated), SCO5314–5320; tetrahydroxynaphthalene (predicted product of the chalcone synthase, which may be further modified by enzymes coded for by other genes in the cluster), SCO1206–1208; isorenieratene, SCO0185–0191; hopanoids (mixture of aminotrihydroxybacteriohopane (shown) and hopene), SCO6759–6771; eicosapentaenoic acid, SCO0124–0129; geosmin, SCO6073; butyrolactones (believed to be assembled by the *scbA* gene product), SCO6266. The structures of the remaining putative secondary metabolites are unknown. The chromosomal location of these clusters and the type of secondary metabolic enzyme(s) coded for are: SCO6429–6438, NRPS; SCO6273–6288 and SCO6826–6827, type I polyketide synthases; SCO7669–7671 and SCO7222, chalcone synthases; SCO5222–5223, sesquiterpene cyclase; SCO5799–5801, siderophore synthetase; SCO1265–1273, type II fatty acid synthase; SCO0381–0401, deoxysugar synthases/ glycosyl transferases.

7683 were deduced to catalyse the biosynthesis of novel siderophores named 'coelichelin'³¹ and 'coelibactin' (G.L.C., unpublished data), respectively. A third cluster, SCO2782–2785, probably directs the biosynthesis of two further siderophores, desferrioxamines G1 and E³². Two large open reading frames (ORFs) (SCO0126 and 0127) code for multi-enzymes with a domain organization very similar to a type I iterative PKS/FAS from a Gram-negative bacterium, *Shewanella* sp., that catalyses biosynthesis of eicosapentaenoic acid³³. We therefore predict a role for this cluster in polyunsaturated fatty acid biosynthesis. Similarly, the cluster SCO6759–6771 has been implicated in hopanoid biosynthesis³⁴, and SCO1206–1208 in tetrahydroxynaphthalene biosynthesis³⁵. The sesquiterpene cyclase coded for by SCO6073 is probably involved in geosmin biosynthesis (B. Gust, K. Fowler, T.K., G.L.C. and K.F.C., personal communication) and SCO0185–0191 probably directs biosynthesis of the carotenoid isorenieratene³⁶.

Although three of the *S. coelicolor* clusters specify antibiotics, most of the others are probably responsible for products with different functions. For example, hopanoids may protect against water loss through the plasma membrane in the aerial mycelium³⁴, and eicosapentaenoic acid may help to maintain membrane fluidity at low temperature. It is notable that at least three clusters probably code for siderophore biosynthesis, implying that *S. coelicolor* is under strong selective pressure to scavenge iron in situations of low iron availability. Thus, products of some of these clusters might accurately be labelled 'stress metabolites', predicted to combat stresses of a physical (desiccation, low temperature), chemical (low iron) or biological (competition) nature.

Cell and developmental biology

Escherichia coli and *B. subtilis* multiply by binary fission, whereas *S. coelicolor* grows as a non-dividing, many-branched mycelium, mainly by tip growth, with multiple copies of the genome in each hyphal compartment. Unigenomic dispersive exospores are borne as chains on specialized, little-branched aerial hyphae that probably extend by intercalary growth³⁷. The genome sequence provides some new insights into this complex life cycle.

Initiation of DNA replication in *S. coelicolor* involves an *oriC*-linked *dnaA* gene, the product of which interacts with an unusually large number (17) of DnaA boxes at the replication origin³⁸. In addition to its initiator function, DnaA is a transcription factor in a diverse range of bacteria³⁹. It is therefore conspicuous that 42 (82%) out of the 51 'strong' DnaA boxes of *S. coelicolor* (TT(G/A)TCCACA³⁸) lie in non-coding DNA upstream of genes. DnaA may conceivably coordinate the replication of multiple genomes in each hyphal compartment with cell-cycle-dependent gene expression.

Our limited understanding of bacterial chromosome partitioning is based largely on studies of low copy number plasmids of Gram-negative bacteria⁴⁰. The *parAB* gene pair on many such plasmids codes for ParA, an ATPase of unclear function, and ParB, which binds one or more *parS* sites near *parA* and *parB*. Many bacteria (including *S. coelicolor*, but not *E. coli*) contain *parAB* genes near *oriC*, and in some cases *parS* target sites have been identified⁴¹. In *S. coelicolor*, there is a high concentration of putative *parS* sites surrounding *oriC*⁴², with 18 'perfect' sites (GTTTCACGTGAAAC) in a 515-kb segment (4,174,551–4,689,985). Unlike DnaA boxes, nearly all of the *parS* sites are immediately downstream of genes, perhaps indicating selection for avoidance of effects on gene expression resulting from ParB–*parS* binding.

Streptomycetes have at least three different kinds of septa⁴³. It is therefore surprising that genes clearly homologous to conserved 'divisome' (cell division) genes of other bacteria are generally present only once or (in the case of *ftsA*) not at all. Presumably the different kinds of cell division involve dedicated accessory proteins. This contrasts with genes coding for enzymes for peptidoglycan synthesis and metabolism: there are eight *ftsI/mrdA* (class

2/3 transpeptidase) and five *mrcA/mrcB* (peptidoglycan synthetase) homologues.

A principal difference between *S. coelicolor* and unicellular rods concerns septum placement. In rods, division involves a centrally located septum, with alternative division sites close to the cell poles usually being silent. This involves the *minC*, *D* and *E* genes in *E. coli*, and the *minC*, *D* and *divIVA* genes of *B. subtilis*⁴⁴. In hyphae of *Streptomyces*, there is no centre point, and division events are usually far from hyphal 'poles'. Consistent with this, there are no *minC*, *minE* or *divIVA*-like genes in *S. coelicolor*. On the other hand, there is a large family of perceptibly *minD*-like genes (which, notably, reveal distant similarity to *parA*). These may control the use of potential division sites at various positions (for example, polar, sub-polar, between pre-existing septa, or at branch points).

Discussion

The genome sequence of *S. coelicolor* has revealed much about the many adaptations of this model actinomycete to life in the highly competitive soil environment. Derived from an ancestor common to other actinomycetes, the chromosome has acquired the ability to replicate in a linear form and appears to have expanded by lateral acquisition and internal duplication of DNA. Chromosome expansion has provided a wealth of genes, allowing the organism a more complex life cycle, adapting to a wider range of environmental conditions and exploiting a greater variety of nutrient sources. This has coincided with an increase in regulatory systems, with a particular emphasis on detection of, and response to, extracellular stimuli. The preferential incorporation (and subsequent maintenance) of occasionally beneficial sequences outside the ancestral core has created chromosome arms comprised mostly of 'non-essential' functions. The abundance of previously uncharacterized metabolic enzymes, particularly those likely to be involved in the production of natural products, is a resource of enormous potential value. Understanding of such enzymes will facilitate the genetic engineering of pathways to produce new compounds with potential therapeutic activity, including much needed antimicrobials⁴⁵. The incomplete genome sequence of an industrial species, *S. avermitilis*⁴⁶, appears to contain a different set of gene clusters for secondary metabolism from *S. coelicolor*. It may be that the arm regions of different streptomycete chromosomes have been accumulated separately, and therefore contain a largely different complement of contingency genes representing a huge pool of metabolic diversity. □

Methods

Genome sequencing

We sequenced the genome of *S. coelicolor* A2(3) from 325 overlapping clones. Of these, 305 were cosmids⁸, one was the terminal plasmid pLUS221 and 19 were selected from a set of 3,456 bacterial artificial chromosomes mapped to the sequences of finished cosmid contigs by end sequencing. The methods for clone growth and isolation, sonication to produce 1.4–2-kb fragments, library preparation in either M13 or pUC18 vectors, and sequencing were as described previously⁴⁷. Most of the clones were digested with *DraI*, and insert purified, before the fragmentation step in order to remove cloning vector. This was not done for those clones known to contain *DraI* sites, and in these cases DNA from the cloning vector was greatly over-represented in the subclone libraries. The finished 325 clones formed a contiguous sequence extending from within the left TIR to the right end of the genome. The genome sequence was completed by extending the incomplete left TIR with a 7-kb consensus sequence copied from the right TIR. The sequence was assembled, finished and annotated as described previously², using Artemis (<http://www.sanger.ac.uk/Software/Artemis>) to collate data and facilitate annotation. Protein families were constructed, independently of annotation, by performing an 'all-against-all' Blast (NCBI Blast version 2) comparison⁴⁸ of proteins within a database containing all predicted protein products from six genomes (Table 2), then single-linkage clustering using a Blast threshold of 70 bits. We checked composition of families using ClustalW⁴⁹. Complex families were resolved by raising the Blast threshold to 100, 150 or 200 bits, as reflected in the hierarchical family numbering system (for example, family 2.1.3 was created using a Blast threshold of 150 on family 2.1).

Received 3 December 2001; accepted 27 March 2002.

1. Hodgson, D. A. Primary metabolism and its control in streptomycetes: a most unusual group of bacteria. *Adv. Microb. Physiol.* **42**, 47–238 (2000).

2. Cole, S. T. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544 (1998).
3. Cole, S. T. *et al.* Massive gene decay in the leprosy bacillus. *Nature* **409**, 1007–1011 (2001).
4. Hopwood, D. A. Forty years of genetics with *Streptomyces*: from *in vivo* through *in vitro* to *in silico*. *Microbiology* **145**, 2183–2202 (1999).
5. Bao, K. & Cohen, S. N. Terminal proteins essential for the replication of linear plasmids and chromosomes in *Streptomyces*. *Genes Dev.* **15**, 1518–1527 (2001).
6. Volf, J. N. & Altenbuchner, J. Genetic instability of the *Streptomyces* chromosome. *Mol. Microbiol.* **27**, 239–246 (1998).
7. Friend, E. J. & Hopwood, D. A. The linkage map of *Streptomyces rimosus*. *J. Gen. Microbiol.* **68**, 187–197 (1971).
8. Redenbach, M. *et al.* A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. *Mol. Microbiol.* **21**, 77–96 (1996).
9. Lobry, J. R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **13**, 660–665 (1996).
10. Eisen, J. A., Heidelberg, J. F., White, O. & Salzberg, S. L. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol.* **1**, (Research) 0011 (2000).
11. Mazodier, P., Thompson, C. & Boccard, F. The chromosomal integration site of the *Streptomyces* element pSAM2 overlaps a putative tRNA gene conserved among actinomycetes. *Mol. Gen. Genet.* **222**, 431–434 (1990).
12. Sezonov, G., Duchène, A. M., Friedmann, A., Guérineau, M. & Pernodet, J. L. Replicase, excisionase, and integrase genes of the *Streptomyces* element pSAM2 constitute an operon positively regulated by the *pra* gene. *J. Bacteriol.* **180**, 3056–3061 (1998).
13. Stover, C. K. *et al.* Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* **406**, 959–964 (2000).
14. Kaneko, T. *et al.* Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res.* **7**, 331–338 (2000).
15. Paget, M. S. B., Hong, H.-J., Bibb, M. J. & Buttner, M. J. in *Control of Bacterial Gene Expression* (eds Hodgson, D. A. & Thomas, C. M.) 105–125 (Cambridge Univ. Press, Cambridge, 2002).
16. Kelemen, G. H. *et al.* A connection between stress and development in the multicellular prokaryote *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **40**, 804–814 (2001).
17. Vohradsky, J. *et al.* Developmental control of stress stimulons in *Streptomyces coelicolor* revealed by statistical analyses of global gene expression patterns. *J. Bacteriol.* **182**, 4979–4986 (2000).
18. Buck, M., Gallegos, M. T., Studholme, D. J., Guo, Y. & Gralla, J. D. The bacterial enhancer-dependent sigma(54) (sigma(N)) transcription factor. *J. Bacteriol.* **182**, 4129–4136 (2000).
19. Berks, B. C., Sargent, F. & Palmer, T. The Tat protein export pathway. *Mol. Microbiol.* **35**, 260–274 (2000).
20. Schneider, D., Bruton, C. J. & Chater, K. F. Duplicated gene clusters suggest an interplay of glycogen and trehalose metabolism during sequential stages of aerial mycelium development in *Streptomyces coelicolor* A3(2). *Mol. Gen. Genet.* **263**, 543–553 (2000).
21. Huang, J., Lih, C. J., Pan, K. H. & Cohen, S. N. Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays. *Genes Dev.* **15**, 3183–3192 (2001).
22. Revill, W. P., Bibb, M. J., Scheu, A. K., Kieser, H. J. & Hopwood, D. A. Beta-ketoacyl acyl carrier protein synthase III (FabH) is essential for fatty acid biosynthesis in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **183**, 3526–3530 (2001).
23. Hopwood, D. A. Genetic contributions to understanding polyketide synthases. *Chem. Rev.* **97**, 2465–2497 (1997).
24. Offner, S., Hofacker, A., Wanner, G. & Pfeifer, F. Eight of fourteen *gvp* genes are sufficient for formation of gas vesicles in halophilic archaea. *J. Bacteriol.* **182**, 4328–4336 (2000).
25. Li, N. & Cannon, M. C. Gas vesicle genes identified in *Bacillus megaterium* and functional expression in *Escherichia coli*. *J. Bacteriol.* **180**, 2450–2458 (1998).
26. Hopwood, D. A., Chater, K. F. & Bibb, M. J. in *Genetics and Biochemistry of Antibiotic Production* (eds Vining, L. C. & Stutterd, C.) 65–102 (Butterworth-Heinemann, Newton, Massachusetts, 1995).
27. Chong, P. P. *et al.* Physical identification of a chromosomal locus encoding biosynthetic genes for the lipopeptide calcium-dependent antibiotic (CDA) of *Streptomyces coelicolor* A3(2). *Microbiology* **144**, 193–199 (1998).
28. Davis, N. K. & Chater, K. F. Spore colour in *Streptomyces coelicolor* A3(2) involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. *Mol. Microbiol.* **4**, 1679–1691 (1990).
29. Challis, G. L., Ravel, J. & Townsend, C. A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7**, 211–224 (2000).
30. Stachelhaus, T., Mootz, H. D. & Marahel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505 (1999).
31. Challis, G. L. & Ravel, J. Coelichelin, a new peptide siderophore encoded by the *Streptomyces coelicolor* genome: structure prediction from the sequence of its non-ribosomal peptide synthetase. *FEMS Microbiol. Lett.* **187**, 111–114 (2000).
32. Imbert, M., Bechet, M. & Blondeau, R. Comparison of the main siderophores produced by some species of *Streptomyces*. *Curr. Microbiol.* **31**, 129–133 (1995).
33. Takeyama, H., Takeda, D., Yazawa, K., Yamada, A. & Matsunaga, T. Expression of the eicosapentaenoic acid synthesis gene cluster from *Shewanella* sp. in a transgenic marine cyanobacterium, *Synechococcus* sp. *Microbiology* **143**, 2725–2731 (1997).
34. Poralla, K., Muth, G. & Hartner, T. Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor* A3(2). *FEMS Microbiol. Lett.* **189**, 93–95 (2000).
35. Funa, N. *et al.* A new pathway for polyketide synthesis in microorganisms. *Nature* **400**, 897–899 (1999).
36. Krügel, H., Krubasik, P., Weber, K., Saluz, H. P. & Sandmann, G. Functional analysis of genes from *Streptomyces griseus* involved in the synthesis of isorenieratene, a carotenoid with aromatic end groups, revealed a novel type of carotenoid desaturase. *Biochim. Biophys. Acta* **1439**, 57–64 (1999).
37. Chater, K. F. & Losick, R. in *Bacteria as Multicellular Organisms* (eds Shapiro, J. A. & Dworkin, M.) 149–182 (Oxford Univ. Press, New York, 1997).
38. Zakrzewska-Czerwinska, J., Jakimowicz, D., Majka, J., Messer, W. & Schrempf, H. Initiation of the *Streptomyces* chromosome replication. *Antonie Van Leeuwenhoek* **78**, 211–221 (2000).
39. Messer, W. & Weigel, C. DnaA initiator—also a transcription factor. *Mol. Microbiol.* **24**, 1–6 (1997).
40. Bignell, C. & Thomas, C. M. The bacterial ParA-ParB partitioning proteins. *J. Biotechnol.* **91**, 1–34 (2001).
41. Lin, D. C. & Grossman, A. D. Identification and characterization of a bacterial chromosome partitioning site. *Cell* **92**, 675–685 (1998).
42. Kim, H. J., Calcutt, M. J., Schmidt, F. J. & Chater, K. F. Partitioning of the linear chromosome during sporulation of *Streptomyces coelicolor* A3(2) involves an *oric*-linked *parAB* locus. *J. Bacteriol.* **182**, 1313–1320 (2000).
43. Kwak, J., Dharmatilake, A. J., Jiang, H. & Kendrick, K. E. Differential regulation of *ftsZ* transcription during septation of *Streptomyces griseus*. *J. Bacteriol.* **183**, 5092–5101 (2001).
44. Jacobs, C. & Shapiro, L. Bacterial cell division: a moveable feast. *Proc. Natl Acad. Sci. USA* **96**, 5891–5893 (1999).
45. Rodriguez, E. & McDaniel, R. Combinatorial biosynthesis of antimicrobials and other natural products. *Curr. Opin. Microbiol.* **4**, 526–534 (2001).
46. Omura, S. *et al.* Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. *Proc. Natl Acad. Sci. USA* **98**, 12215–12220 (2001).
47. Harris, D. & Murphy, L. in *Methods in Molecular Biology* (eds Starkey, M. P. & Elsaswarapu, R.) Vol. 175, 217–234 (Humana, Totowa, New Jersey, 2001).
48. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
49. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
50. Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85**, 2444–2448 (1988).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We would like to acknowledge the support of the Wellcome Trust Sanger Institute core sequencing and informatics groups. This work was funded by the Biotechnology and Biological Sciences Research Council and by the Wellcome Trust through its Beowulf Genomics Initiative. C.W.C. and C.-H.H. were supported by the R.O.C. National Science Council and Ministry of Education.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.D.B. (e-mail: sdb@sanger.ac.uk) or D.A.H. (e-mail: david.hopwood@bbsrc.ac.uk). The complete sequence is deposited in GenBank/EMBL under accession number AL645882.

Identification of iron sulphide grains in protoplanetary disks

L. P. Keller*, S. Hony†, J. P. Bradley‡, F. J. Molster§, L. B. F. M. Waters†☆, J. Bouwman**, A. de Koter†, D. E. Brownlee||, G. J. Flynn¶, T. Henning# & H. Mutschke#

* Mail Code SR, NASA Johnson Space Center, Houston, Texas 77058, USA

† Astronomical Institute Anton Pannekoek, Amsterdam, Kruislaan 403, NL-1098 SJ, The Netherlands

‡ School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA

§ ESTEC/ESA, RSSD-ST, Keplerlaan 1, NL-2201 AZ Noordwijk, The Netherlands

|| Department of Astronomy, University of Washington, Seattle, Washington 98195, USA

¶ Department of Physics, SUNY, Plattsburgh, New York 12901, USA

Astrophysical Institute and University Observatory, Jena, Germany

☆ Instituut voor Sterrenkunde, Katholieke Universiteit Leuven, Celestijnenlaan 200B, B-3001 Heverlee, Belgium

** CEA, DSM, DAPNIA, Service d'Astrophysique, CEN Saclay, F-91191 Gif-sur-Yvette Cedex, France

Sulphur is depleted in cold dense molecular clouds with embedded young stellar objects¹, indicating that most of it probably resides in solid grains. Iron sulphide grains are the main sulphur species in cometary dust particles^{2,3}, but there has been no direct evidence for FeS in astronomical sources⁴, which poses a considerable problem, because sulphur is a cosmically abundant element. Here we report laboratory infrared spectra of FeS grains from primitive meteorites, as well as from pyrrhotite ($[\text{Fe}, \text{Ni}]_{1-x}\text{S}$) grains in interplanetary dust, which show a broad FeS feature centred at ~ 23.5 micrometres. A similar broad feature is seen in the infrared spectra of young stellar objects, implying that FeS grains are an important but previously unrecognized component of circumstellar dust. The feature had previously been attributed to FeO⁵⁻⁷. The observed astronomical line strengths are generally consistent with the depletion of sulphur from the gas phase¹, and with the average Galactic sulphur/silicon abundance ratio⁸. We conclude that the missing sulphur has been found.

Spectroscopic analyses of primitive meteorites, cometary dust particles and synthetic analogue materials in the laboratory provide constraints and serve as 'ground truth' for evaluating various hypotheses on the nature of interstellar grains. Comets are primitive bodies that are widely believed to be a large reservoir of preserved interstellar and circumstellar grains along with presolar organic matter. The comparison of laboratory infrared data on cometary dust with astronomical spectra has been facilitated by the high-quality spectra obtained from the Infrared Space Observatory (ISO)⁹. An emission feature is observed^{5,10} in ISO spectra at ~ 23.5 μm and is compared here with iron sulphide minerals from meteorites, interplanetary dust particles (IDPs) and terrestrial sources which show infrared features in the same spectral range (Fig. 1). Figure 2 shows infrared spectra from iron oxide and iron sulphide mineral standards. Wüstite (FeO) and magnetite (Fe_3O_4) show only one strong, narrow feature at ~ 17.5 μm . FeO has been used to model the 23- μm feature in ISO spectra because the position of the FeO feature can be shifted to 23 μm in calculated spectra using experimentally derived optical constants combined with theoretical grain shapes and sizes⁷. Pyrrhotite exhibits a strong, broad, asymmetric absorption maximum centred at ~ 23.5 μm , whereas the feature in troilite is closer to 22 μm . The pyrrhotite spectrum also shows a weak shoulder at ~ 17.5 μm that is consistent with iron oxide. Pyrrhotite is well known to quickly form a thin surface oxide layer when exposed to ambient atmospheric conditions. Figure 3 shows infrared spectra from two pyrrhotite-rich IDPs (L2011*B6 and

U2012A-2J) together with 23.5- μm features from two Herbig stars (HD163296 and AB Aurigae)¹¹. The sulphide mineral standards, as well as the sulphides in the IDPs, provide an excellent match to the 23.5- μm feature in young stellar objects in terms of peak positions, shapes and widths (Figs 1 and 3), although minor differences do exist.

The 23.5- μm band observed in circumstellar disks surrounding young stars matches that of troilite (FeS) in the evolved star M2-43 (Fig. 1). The identification of FeS in M2-43 is based on the good match in peak positions between laboratory and astronomical spectra of not only the 23.5- μm band, but also the longer-wavelength bands at 34-, 38- and 44- μm , and because iron sulphides are a predicted dust component in carbon-rich evolved stars¹². We observe that the 23.5- μm band is consistently much stronger in the astronomical data than in the laboratory spectrum of FeS derived from optical constants¹³. The 34-, 38- and 44- μm bands

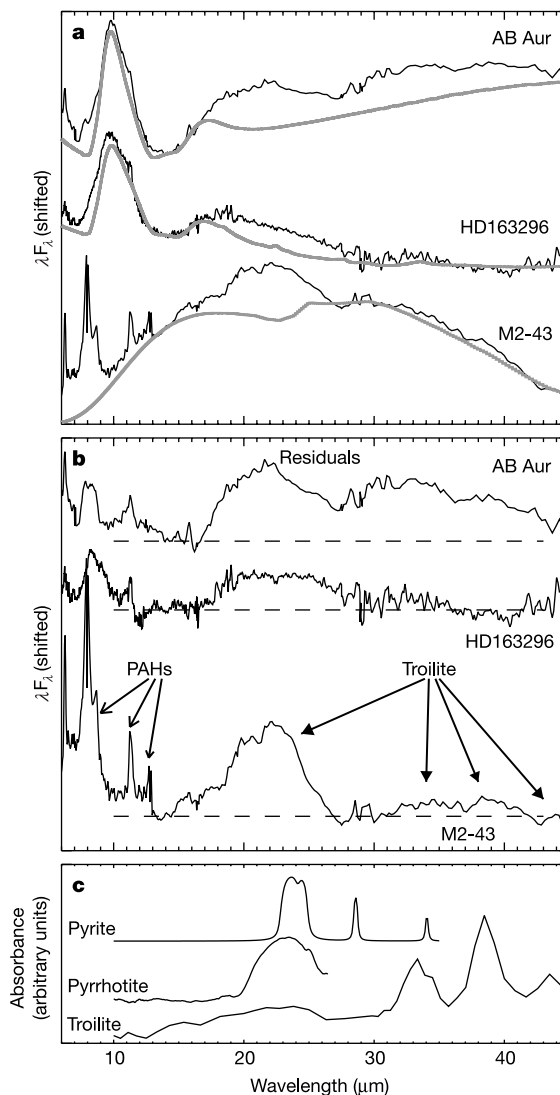


Figure 1 Infrared spectra of young stars showing a pronounced 23.5- μm feature from iron sulphide grains. **a**, The mid-infrared spectra of the young stars AB Aurigae, HD163296 and the evolved star M2-43. The grey lines in the AB Aur and HD163296 spectrum represent model fits to the full ISO spectra using amorphous silicates, metallic Fe and carbonaceous material¹⁰. The grey line in the M2-43 spectrum is a fit to the continuum using amorphous carbon and MgS. **b**, Residuals after subtracting the models from the observed ISO spectra. **c**, Infrared spectra of troilite (FeS) and of pyrite (FeS₂) calculated from optical constants, and of pyrrhotite (Fe_{1-x}S) from laboratory measurements. We note the major resonances of FeS at 23, 33 and 38 μm in the residual spectrum of M2-43. PAHs, polycyclic aromatic hydrocarbons; λF_{λ} , flux in units of W m^{-2} .

of troilite, although weakly present in the evolved star, are not detected in the young stars; however, we point out that there is spectral structure present in the AB Aurigae spectrum (at low spectral contrast) at approximately the correct wavelength positions for the long-wavelength peaks of FeS (Fig. 1). The lack of a clear signature of FeS in the 30- μm region of the young star can be attributed largely to temperature differences between the FeS in the evolved star and young star. All published spectra of iron sulphides show a characteristic 23.5- μm feature and additional features at longer wavelengths ($>30\mu\text{m}$)^{13–15}. The position and strength of these long-wavelength features vary with mineralogical speciation (Fig. 1) and crystallinity¹⁴. In addition, grain size and grain shape have significant effects on the sharpness and position of the infrared features. The astronomical spectra almost certainly reflect a mixture of iron sulphide phases (as do most cometary IDPs³), which also serves to enhance the 23.5- μm band strength with respect to the longer-wavelength bands.

If the ISO feature is indeed due to iron sulphides, then the grains must be small because we were only able to collect useful absorption spectra from specimens $<200\text{ nm}$ thick. When the sample thickness exceeds $\sim 200\text{ nm}$, the sulphide standards became infrared opaque, which is consistent with our calculated mass absorption coefficient. There is an extensive literature on similar effects observed for SiC and other dense, strongly infrared-absorbing materials^{15,16}. ISO spectra from comet Hale–Bopp show a significantly sharper feature at $\sim 23.5\mu\text{m}$ attributed to forsterite (Mg_2SiO_4)^{17,18}. The overlap with forsterite makes it difficult to determine if a broader underlying feature, consistent with iron sulphide, is present in the Hale–Bopp spectrum. As noted above, however, the grain size of the sulphides strongly influences whether or not the feature will be detected.

One fundamental question regarding these results is whether

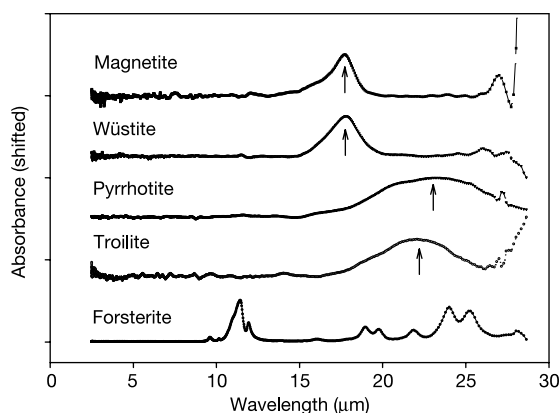


Figure 2 Infrared transmission spectra from iron oxides are compared with iron sulphide standards. The iron oxides display a narrow diagnostic feature between 17 and 18 μm whereas the iron sulphides have a broad feature showing its intensity maximum between 22 and 24 μm . Also shown for comparison is the spectrum from meteoritic forsterite (Mg_2SiO_4) which displays several Si–O bending modes between 15 and 25 μm . Spectra were obtained using a SpectraTech I μ s Fourier transform infrared microspectrometer on the intense infrared beamline (U10B) at the National Synchrotron Light Source at Brookhaven National Laboratory. The infrared instrument is sensitive over the 2.5–28.5 μm (4,000–350 cm^{-1}) wavelength range, although the spectra become very noisy above $\sim 26\mu\text{m}$ owing to absorption of the signal in the KBr beamsplitter. Infrared spectra were collected from ultramicrotome thin sections of terrestrial pyrrhotite, troilite, synthetic wüstite and magnetite that were sectioned in epoxy and supported on carbon-film transmission electron microscope (TEM) grids. We used an aperture of $32 \times 32\mu\text{m}$ for the measurements because smaller apertures resulted in spectral artefacts from diffraction effects at long wavelengths. Final baseline corrected (and smoothed) spectra were obtained by subtracting the background spectrum from the thin carbon film substrates from the sample spectrum. Between 256 and 1,000 scans (interferograms) were collected from each specimen with 4 cm^{-1} resolution.

the amount of iron sulphide required to produce the infrared feature is consistent with abundance constraints imposed by gas-phase depletions in dense molecular clouds. For this analysis, we compared the intrinsic strength of the 18- μm silicate feature to the 23.5- μm sulphide feature, because it minimizes the effects of temperature gradients on the band-strength analysis (the 10- μm silicate band is very sensitive to the presence of hot silicate grains). The relative band strength can be related quantitatively to composition provided we know μ^{sil} and μ^{FeS} (the molecular weights of astronomical silicate and FeS), $\kappa_{\text{sil}}^{18\mu\text{m}}$ and $\kappa_{\text{FeS}}^{23.5\mu\text{m}}$ (the mass absorption coefficients for the silicate and FeS respectively), and $I_{18\mu\text{m}}$ and $I_{23.5\mu\text{m}}$ (the intensities of the 18- and 23.5- μm features observed in the ISO spectra). There are however, several assumptions that enter into this calculation. We assume that the grain temperature differences are negligible, and we know that the full-widths at half-maximum for both features are approximately the same, so these terms drop out of the calculation. The mass absorption coefficient of the 18- μm silicate feature depends on the composition of the silicate¹⁹. We adopt a $\kappa_{\text{sil}}^{18\mu\text{m}}$ of $\sim 1,000\text{ cm}^2\text{ g}^{-1}$, which is somewhat lower than the widely used value²⁰ of $1,200\text{ cm}^2\text{ g}^{-1}$, because we do not factor in the underlying continuum. We have calculated a $\kappa_{\text{FeS}}^{23.5\mu\text{m}}$ of $\sim 1,340\text{ cm}^2\text{ g}^{-1}$ from our laboratory spectra of pyrrhotite, which is similar to other reported values¹⁴, but is significantly larger than that obtained using reflectance methods¹³. This large difference in cross-sections requires further investigation. Using the band

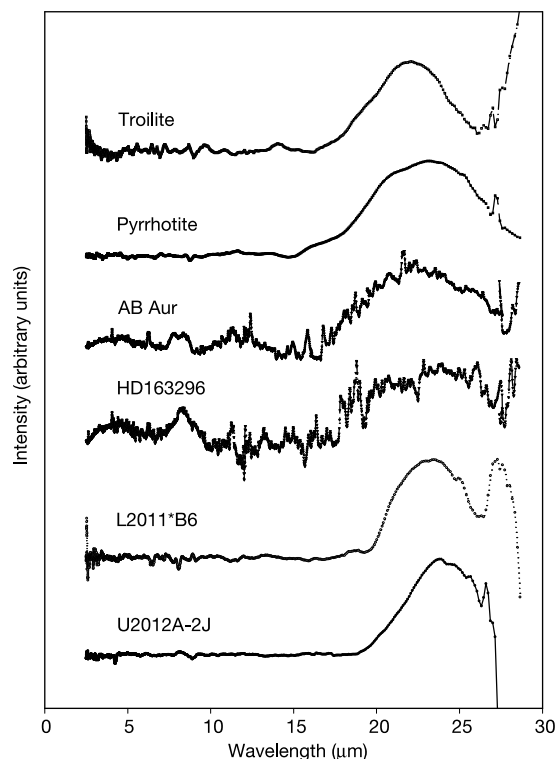


Figure 3 Infrared spectra of sulphide standards and sulphide-rich IDPs compared to ‘23.5- μm ’ features in two Herbig stars showing similar peak positions, shapes, and bandwidths. The ‘23.5- μm ’ features of AB Aur and HD 163296 were obtained by subtracting our best-fit radiative transfer model¹⁰ without the contribution of ‘FeO’ (see ref. 11). The dust components used in our best-fit model include glassy silicates, forsterite, carbonaceous material, metallic iron and water ice. The IDP infrared spectra were collected from ultramicrotome thin sections ($\sim 70\text{ nm}$ thick) of IDPs that had been sectioned in sulphur and supported on carbon-film TEM grids under the same conditions as given in Fig. 2. L2011*B6 is a fragment of a deuterium-rich cluster IDP, indicating that it contains some presolar molecular cloud material. The mineralogy of L2011*B6 is dominated by pyrrhotite grains, as well as minor GEMS (glass with embedded metal and sulphides), enstatite and carbonaceous material. U2012A-2J is also dominated by low-Ni Fe-sulphides.

strengths extracted from the ISO spectra of HD163296, we obtain a $[S/Si]_{\text{calc}} \cong 0.63$. This ratio is consistent with the cosmochemical (solar) abundance⁸ of Si and S (atom ratio 0.52). We conclude from this analysis that the sulphur abundance constraints are broadly consistent with the infrared spectra, although the error bars are large. The implication of this analysis is that most if not all of the sulphur resides in solid grains of FeS.

In the diffuse interstellar medium, sulphur is usually not strongly depleted from the gas phase²¹, indicating that little sulphur is incorporated into solid grains in this environment. Either iron sulphides are not produced in significant quantities in stellar outflows, or their lifetime in the interstellar medium is very short owing to rapid destruction. One probable mechanism for their destruction is the reduction of iron sulphide through irradiation, a process that produces Fe metal as a by-product and returns sulphur to the gas phase. A similar mechanism is known to occur in lunar Fe-oxides exposed to the solar wind²² as well as in GEMS (proposed presolar silicate grains) in cometary IDPs²³. The situation is very different in cold, dense molecular clouds, where sulphur is highly depleted from the gas phase¹. In these environments, only trace amounts of sulphur in molecular form or in ice mantles are detected⁴. Therefore, we believe the bulk of the depletion must be accommodated through incorporation of sulphur into solid grains of iron sulphide. The bulk of the iron sulphide is formed either in the collapse phase of the molecular cloud, or in the disk surrounding the young star. Support for this hypothesis comes from observations of iron sulphide grains in IDPs which are intimately associated with carbonaceous material showing H and N isotopic signatures of molecular cloud environments^{24,25}.

From ISO data, crystalline magnesium-rich silicates (forsterite and enstatite) have been recently discovered in a number of sources. It is probable that iron sulphides are manifest in many ISO spectra as well. Iron sulphides, magnesium-rich crystalline silicates and amorphous silicates are the main components of primitive cometary IDPs. These results thus provide new links between primitive IDPs, comets and presolar materials, and resolves a long-standing dilemma. Iron sulphides have long been recognized as a major component of primitive meteoritic materials, but until now they had escaped detection in astronomical observations of objects similar to the early solar nebula. Isotopic measurements for both iron and sulphur in primitive IDPs may uniquely identify presolar iron sulphide grains. A major implication of this work is that early Earth and other Solar System bodies accreted most of their sulphur in the form of crystalline iron sulphide grains. □

Received 30 November 2001; accepted 5 March 2002.

- Joseph, C. L., Snow, T. E., Seab, C. G. & Crutcher, R. M. Interstellar abundances in dense, moderately reddened lines of sight. I. Observational evidence for density-dependent depletion. *Astrophys. J.* **309**, 771–782 (1986).
- Schulze, H., Kissel, J. & Jessberger, E. K. in *Stardust to Planetesimals* (eds Pendleton, T. L. & Tielens, A. G. G. M.) *Astron. Soc. Pacific. Conf. Ser.* **122**, 397–414 (1997).
- Dai, Z. R. & Bradley, J. P. Iron-nickel sulfides in anhydrous interplanetary dust particles. *Geochim. Cosmochim. Acta* **65**, 3601–3612 (2001).
- Palumbo, M. E., Geballe, T. R. & Tielens, A. G. G. M. Solid carbonyl sulfide (OCS) in dense molecular clouds. *Astrophys. J.* **479**, 839–844 (1997).
- Meeus, G. *et al.* ISO spectroscopy of circumstellar dust in 14 Herbig Ae/Be systems: Towards an understanding of dust processing. *Astron. Astrophys.* **365**, 476–490 (2001).
- Waters, L. B. F. M., Molster, F. J. & Waelkens, C. in *Solid Interstellar Matter: The ISO Revolution* (eds D'Hendecourt, L., Joblin, C. & Jones, A.) 219–229 (Springer, Berlin, 1998).
- Henning, T., Begemann, B., Mutschke, H. & Dorschner, J. Optical properties of oxide dust grains. *Astron. Astrophys. Suppl.* **112**, 143–149 (1995).
- Anders, E. & Grevesse, N. Abundances of the elements—Meteoritic and solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1988).
- Kessler, M. F. *et al.* The Infrared Space Observatory (ISO) mission. *Astron. Astrophys.* **315**, L27–L31 (1996).
- Bouwman, J., de Koter, A., van den Anker, M. E. & Waters, L. B. F. M. The composition of the circumstellar dust around the Herbig Ae stars AB Aur and HD 163296. *Astron. Astrophys.* **360**, 213–226 (2000).
- van den Anker, M. E. *et al.* ISO Spectroscopy of circumstellar dust in the Herbig Ae systems AB Aur and HD 163296. *Astron. Astrophys.* **357**, 325–329 (2000).
- Lodders, K. & Fegley, B. in *Asymptotic Giant Branch Stars* (eds LeBrete, T., Lebre, A. & Waelkens, C.) 279–293 (IAU Symposium 191, 1999).

- Begemann, B. *et al.* A laboratory approach to the interstellar sulfide dust problem. *Astrophys. J.* **423**, L71–L74 (1994).
- Nuth, J. A. *et al.* Laboratory infrared spectra of predicted condensates in carbon-rich stars. *Astrophys. J.* **290**, L41–L43 (1985).
- Lennie, A. R. & Vaughan, D. J. Kinetics of the marcasite-pyrite transformation: An infrared spectroscopic study. *Am. Mineral.* **77**, 1166–1177 (1992).
- Hofmeister, A. M., Rosen, L. J. & Speck, A. K. in *Thermal Emission Spectroscopy and Analysis of Dust, Disks, and Regoliths* (eds Sitko, M. L., Sprague, A. L. & Lynch, D. K.) *Astron. Soc. Pacific. Conf. Ser.* **196**, 291–300 (2000).
- Malfait, K. *et al.* The spectrum of young star HD 100546 observed with the Infrared Space Observatory. *Astron. Astrophys.* **332**, L25–L28 (1998).
- Crovisier, J. *et al.* The spectrum of comet Hale-Bopp (C/1995 O1) observed with the Infrared Space Observatory at 2.9 astronomical units from the Sun. *Science* **275**, 1904–1907 (1997).
- Mutschke, H. *et al.* Steps toward interstellar silicate mineralogy. III. The role of aluminum in circumstellar amorphous silicates. *Astron. Astrophys.* **333**, 188–199 (1998).
- Draine, B. T. & Lee, H. M. Optical properties of interstellar graphite and silicate grains. *Astrophys. J.* **285**, 89–108 (1984).
- Savage, B. D. & Sembach, K. R. Interstellar abundances from absorption-line observations with the Hubble Space Telescope. *Annu. Rev. Astron. Astrophys.* **34**, 279–330 (1996).
- Christoffersen, R., Keller, L. P. & McKay, D. S. Microstructure, chemistry, and origin of grain rims on ilmenite from the lunar soil finest fraction. *Meteorit. Planet. Sci.* **31**, 835–848 (1996).
- Bradley, J. P. Chemically anomalous preaccretorally irradiated grains in interplanetary dust from comets. *Science* **265**, 925–929 (1994).
- Keller, L. P., Messenger, S. & Bradley, J. P. Analysis of a deuterium-rich interplanetary dust particle and implications for presolar materials in IDPs. *J. Geophys. Res.* **A 105**, 10397–10402 (2000).
- Keller, L. P., Messenger, S., Miller, M. A. & Thomas, K. L. Nitrogen speciation in a ¹⁵N-enriched interplanetary dust particle. *Lunar Planet. Sci.* **28**, 1811–1812 (1997).

Acknowledgements

We thank L. Miller, L. Carr and G. Williams at the National Synchrotron Light Source at Brookhaven National Laboratory for technical assistance. Much of this work was performed while L.P.K. was a Senior Research Scientist at MVA, Inc. F.J.M. acknowledges support from an NWO Talent Grant. L.B.F.M.W., S.H., A.d.K. and J.B. acknowledge financial support from an NWO Pioneer Grant. We thank A. Li and D. Wooden for comments on the manuscript. This work was supported in part by NASA and ESA.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.P.K. (e-mail: lindsay.p.keller@jsc.nasa.gov).

Formation and propagation of matter-wave soliton trains

Kevin E. Strecker*, Guthrie B. Partridge*, Andrew G. Truscott*† & Randall G. Hulet*

* Department of Physics and Astronomy and Rice Quantum Institute, Rice University, Houston, Texas 77251, USA

Attraction between the atoms of a Bose–Einstein condensate renders it unstable to collapse, although a condensate with a limited number of atoms¹ can be stabilized² by confinement in an atom trap. However, beyond this number the condensate collapses^{3–5}. Condensates constrained to one-dimensional motion with attractive interactions are predicted to form stable solitons, in which the attractive forces exactly compensate for wave-packet dispersion¹. Here we report the formation of bright solitons of ⁷Li atoms in a quasi-one-dimensional optical trap, by magnetically tuning the interactions in a stable Bose–Einstein condensate from repulsive to attractive. The solitons are set in motion by offsetting the optical potential, and are observed to propagate in the potential for many oscillatory cycles without spreading. We observe a soliton train, containing many solitons; repulsive interactions between neighbouring solitons are inferred from their motion.

† Present address: Research School of Physical Sciences and Engineering, Australian National University, Canberra, ACT 0200, Australia.

Dispersion and diffraction cause localized wave packets to spread as they propagate. Solitons may be formed when a nonlinear interaction produces a self-focusing of the wave packet that compensates for dispersion. Such localized structures have been observed in many physical systems including water waves, plasma waves, sound waves in liquid helium, particle physics, and in optics⁶. A Bose–Einstein condensate can be described by the nonlinear Schrödinger equation, for which the interaction term is cubic in the condensate wavefunction⁷. For attractive interactions, this equation has the same form as the equation for an optical wave propagating in a medium with a cubic, self-focusing (Kerr) nonlinearity and, in this sense, bright matter-wave solitons in one dimension are similar to optical solitons in optical fibres. Dark solitons have been recently studied in condensates with repulsive atomic interactions^{8–10}, but they are limited in that they can only exist within the condensate itself. Bright solitons, on the other hand, may propagate over much larger distances, and are themselves condensates. A similar experiment to ours has recently been performed by L. Khaykovich *et al.* (personal communication).

The degree of radial confinement necessary to achieve soliton stability has been investigated theoretically^{11–14}. Assuming cylindrically symmetric harmonic confinement with axial and radial oscillation frequencies of ω_z and ω_r , respectively, radial excitations are suppressed in the so-called quasi-one-dimensional (quasi-1D) regime, where $\hbar\omega_r$ exceeds the magnitude of the mean-field interaction energy. This requirement is equivalent to a limitation on the condensate occupation number of $N < l_r/|a|$, where $l_r = (\hbar/m\omega_r)^{1/2}$ is the radial scale length, m is the atomic mass and a is the s -wave scattering length characterizing the two-body interactions. The interactions are effectively attractive for $a < 0$ and repulsive for $a > 0$. Achievement of the quasi-1D regime has been recently demonstrated in the case of ⁷Li (ref. 15) and ²³Na (ref. 16) condensates.

The apparatus for producing a quantum degenerate gas of ⁷Li atoms was described previously¹⁷, although a new magnetic trap with axial and radial frequencies of 70 Hz and 800 Hz, respectively, has been incorporated. Atoms in the $(F, m_F) = (2, 2)$ sublevel, where F and m_F are, respectively, the total electronic angular momentum and its projection, are evaporatively cooled in the magnetic trap to a temperature of $\sim 1 \mu\text{K}$. Atoms in the $(1, 1)$ state, which are not magnetically trappable, are needed in the final stages of the experiment, so the $(2, 2)$ atoms are transferred to an optical trap. The optical trap consists of a single, focused, red-detuned laser beam propagating in the axial direction for radial confinement, and a separated pair of cylindrically focused blue-detuned laser beams ('end caps') propagating in the radial plane, for axial confinement. The single beam is provided by an infrared Nd:YAG laser, with a

wavelength of 1,064 nm, focused to a $1/e^2$ intensity radius of $47 \mu\text{m}$ and with a power of up to 750 mW. The radial confinement potential is approximately harmonic, and at the highest power, matches well with the magnetic trap potential. This single axial beam provides a very weak axial restoring potential, which is also approximately harmonic, with a frequency of ~ 4 Hz for oscillation amplitudes less than the Rayleigh length of 6.5 mm. The end caps are generated from the second harmonic of another Nd:YAG laser, have $1/e^2$ radii of $22 \mu\text{m}$ axially and $100 \mu\text{m}$ radially, a power of 350 mW in each beam, and are separated by $230 \mu\text{m}$. The end caps create a box-like potential in the axial direction, which helps to better match the magnetic trap potential. Once the optical trap lasers are switched on, the magnetic trap is switched off. After 50 ms, a bias field of up to 1,000 G is applied. A period of ~ 200 ms is allowed for the bias field to stabilize at the chosen value before the atoms are transferred from the $(2, 2)$ state to the $(1, 1)$ state by an adiabatic microwave sweep of 15 ms in duration and 1 MHz in width. The purity of the $(1, 1)$ state population is measured to be greater than 98%.

Attractive interactions between atoms in the $(2, 2)$ state limit the number of possible condensate atoms to a very small fraction of the total number. However, interactions between $(1, 1)$ state atoms at zero magnetic field are repulsive with a scattering length $a = 5a_0$, where a_0 is the Bohr radius¹⁸. Although this fact ensures the stability of the condensate, the rate of its formation is limited by the rate of thermalization, which scales as a^2 . In order to increase this rate, a magnetically tuned Feshbach resonance^{19,20} is used to increase the magnitude of a . Figure 1 shows the results of a calculation of a versus the magnetic field B , which exhibits a collisional resonance near 725 G. The rate of loss of atoms is observed to increase rapidly near the resonance (Fig. 2), in accordance with previous investigations^{20,21}. Atoms are evaporatively cooled in the optical trap by halving the intensity of the infrared beam, and by tuning the bias field to 710 G, where the scattering length is large ($\sim 200a_0$) and positive. Absorption imaging indicates that a condensate is formed with up to $\sim 3 \times 10^5$ atoms. Our techniques are similar to those used to make condensates of ⁸⁵Rb, where a Feshbach resonance was used to manipulate the sign and magnitude of a (ref. 21).

Following the formation of the condensate at large positive a , the field is ramped down as $e^{-t/\tau}$ (where t is time and $\tau = 40$ ms), to a selected field between 545 G and 630 G, where a is small and negative or small and positive (Fig. 1). The condensate can be created on the side of the optical potential by axially displacing the focus of the infrared beam relative to the centres of the magnetic trap and the box potential formed by the end caps. The end caps

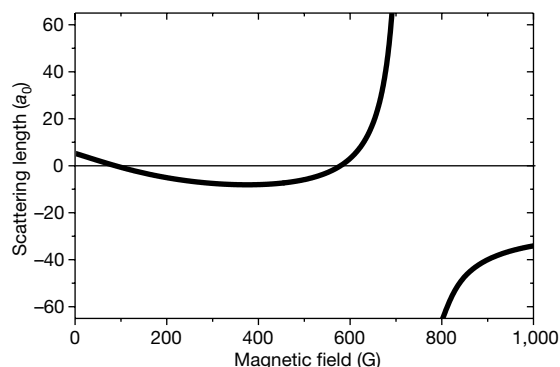


Figure 1 Feshbach resonance. Calculation of the scattering length versus magnetic field for atoms in the $(1, 1)$ state of ⁷Li using the coupled channels method¹⁸. The field axis has been scaled here by a factor of 0.91, to agree with the measured resonance position of 725 G shown in Fig. 2. The scattering length is given in units of the Bohr radius, a_0 .

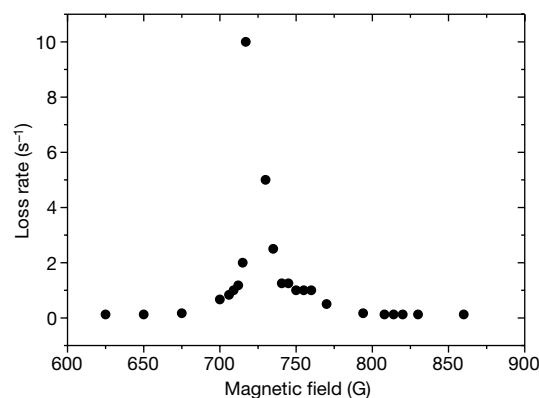


Figure 2 Measured rate of inelastic collisional loss of atoms near the Feshbach resonance. The temperature is $\sim 1 \mu\text{K}$, which is above the transition temperature for Bose–Einstein condensation. The initial peak density is estimated to be $\sim 6 \times 10^{12} \text{ cm}^{-3}$. The rate of loss is given by the time for the number of trapped atoms to fall to e^{-1} of the initial number. The magnetic field is determined spectroscopically by measuring the frequency of the $(2, 2) \rightarrow (1, 1)$ transition to within an uncertainty of 0.1 G.

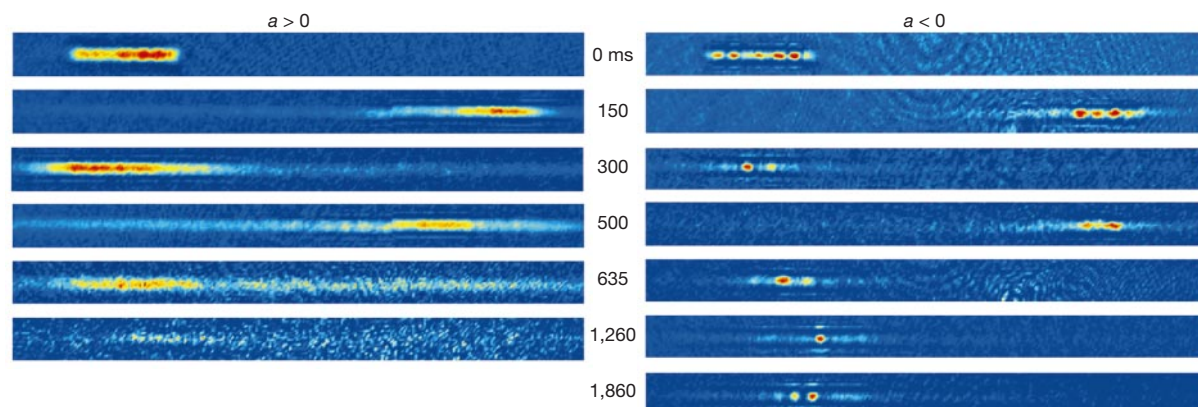


Figure 3 Comparison of the propagation of repulsive condensates with atomic solitons. The images are obtained using destructive absorption imaging, with a probe laser detuned 27 MHz from resonance. The magnetic field is reduced to the desired value before switching off the end caps (see text). The times given are the intervals between turning off the end caps and probing (the end caps are on for the $t = 0$ images). The axial dimension of each image frame corresponds to 1.28 mm at the plane of the atoms. The amplitude of

oscillation is $\sim 370 \mu\text{m}$ and the period is 310 ms. The $a > 0$ data correspond to 630 G, for which $a \approx 10a_0$, and the initial condensate number is $\sim 3 \times 10^5$. The $a < 0$ data correspond to 547 G, for which $a \approx -3a_0$. The largest soliton signals correspond to $\sim 5,000$ atoms per soliton, although significant image distortion limits the precision of number measurement. The spatial resolution of $\sim 10 \mu\text{m}$ is significantly greater than the expected transverse dimension $l_r \approx 1.5 \mu\text{m}$.

prevent the condensate from moving under the influence of the infrared potential until, at a certain instant, the end caps are switched off and the condensate is set in motion. The condensate is allowed to evolve for a set period of time before an image is taken. As shown in Fig. 3, the condensate spreads for $a > 0$, while for $a < 0$, non-spreading, localized structures (solitons) are formed. Solitons have been observed for times exceeding 3 s, a limitation that we believe is due to loss of atoms rather than wave-packet spreading.

Multiple solitons ('soliton trains') are usually observed, as is evident in Figs 3 and 4. We find that typically four solitons are created from an initially stationary condensate. Although multi-soliton states with alternating phase are known to be stationary states of the nonlinear Schrödinger equation^{14,22,23}, mechanisms for their formation are diverse. It was proposed that a soliton train could be generated by a modulational instability²⁴, where in the case of a condensate, phase fluctuations would produce a maximum rate of amplitude growth at a wavelength approximately equal to the condensate healing length $\xi = (8\pi n|a|)^{-1/2}$, where n is the atomic density²³. As a and ξ are dynamically changing in the experiment, the expected number of solitons N_s is not readily estimated from a static model. Experimentally, we detect no

significant difference in N_s when the time constant, τ , for changing the magnetic field is varied from 25 ms to 200 ms. We investigated the dependence of N_s on condensate velocity v by varying the interval Δt between the time the end caps are switched off to the time when a changes sign. We find that N_s increases linearly with Δt , from ~ 4 at $\Delta t = 0$ to ~ 10 at $\Delta t = 35$ ms. As the axial oscillation period is ~ 310 ms, $v \propto \Delta t$ in the range of Δt investigated.

The alternating phase structure of the soliton train can be inferred from the relative motion of the solitons. Non-interacting solitons, simultaneously released from different points in a harmonic potential, would be expected to pass through one another. But this is not observed, as can be seen from Fig. 4, which shows that the spacing between the solitons increases near the centre of oscillation and bunches at the end points. This is evidence of a short-range repulsive force between the solitons. Interaction forces between solitons have been found to vary exponentially with the distance between them, and to be attractive or repulsive depending on their relative phase²⁵. Because of the effect of wave interference on the kinetic energy, solitons that differ in phase by π will repel, while those that have the same phase will attract. An alternating phase structure can be generated in the initial condensate by a phase

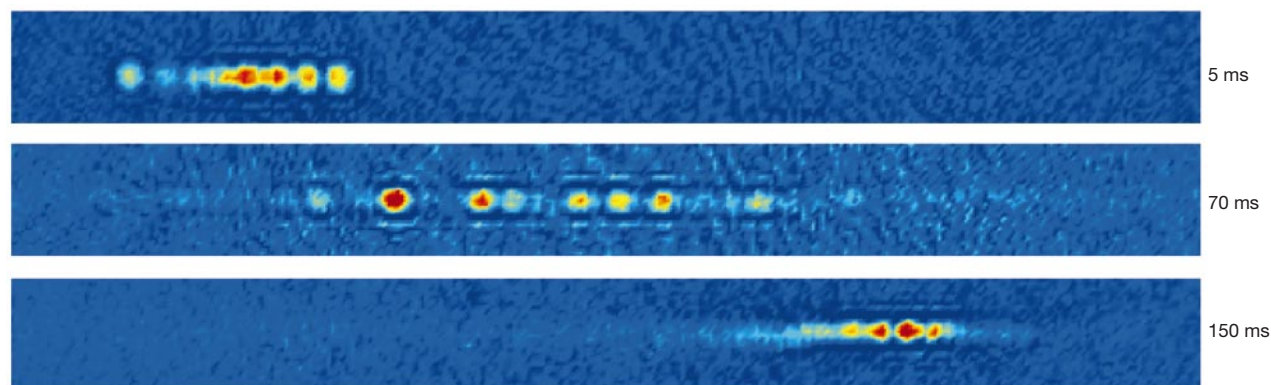


Figure 4 Repulsive interactions between solitons. The three images show a soliton train near the two turning points and near the centre of oscillation. The spacing between solitons is compressed at the turning points, and spread out at the centre of the oscillation. A simple model based on strong, short-range, repulsive forces between nearest-neighbour solitons indicates that the separation between solitons oscillates at approximately twice the trap frequency, in agreement with observations. The number of

solitons varies from image to image because of shot to shot experimental variations, and because of a very slow loss of soliton signal with time. As the axial length of a soliton is expected to vary as $1/N$ (ref. 11), solitons with small numbers of atoms produce particularly weak absorption signals, scaling as N^2 . Trains with missing solitons are frequently observed, but it is not clear whether this is because of a slow loss of atoms, or because of sudden loss of an individual soliton.

gradient, $d\phi/dz$, across the condensate. Such a gradient may be imprinted by a condensate velocity, because $d\phi/dz = mv/\hbar$, where m is the atomic mass. If N_s is identified with ϕ/π , the model predicts $N_s \propto v$, in agreement with the observed v -dependent part of N_s . Furthermore, for the largest v and for parameters consistent with the experiment, the model gives $N_s \approx 15$, in rough agreement with observation.

For a soliton with $a = -3a_0$, the calculated maximum number of atoms that ensures stability is only $\sim 6,000$ per soliton^{11,13}, which accounts for far fewer atoms than the number contained in the initial repulsive condensate. Apparently, most of the atoms from the collapsing condensate are lost, while only a small fraction remain as solitons. Immediately after switching a from positive to negative, we observe a diffuse background of atoms spreading out axially. This observation is reminiscent of the jet emitted by a ⁸⁵Rb condensate after switching from repulsive to attractive interactions⁵. In our system, which is in the quasi-1D regime, the remnant atoms form solitons with atom number near their stability limit.

The remarkable similarities between bright matter-wave solitons and optical solitons in fibres²⁶ emphasize the intimate connection between atom optics with Bose-Einstein condensates²⁷ and light optics. Many issues remain to be addressed, however, including the dynamical process of soliton formation. In addition, further investigation of soliton interactions and collisions could be undertaken with this system. Finally, we speculate that an 'atomic soliton laser', based on bright matter-wave solitons, may prove useful for precision measurement applications, such as atom interferometry²⁸. □

Received 19 March; accepted 18 April 2002.

Published online 1 May 2002; DOI 10.1038/nature747.

- Ruprecht, P. A., Holland, M. J., Burnett, K. & Edwards, M. Time-dependent solution of the nonlinear Schrödinger equation for Bose-condensed trapped neutral atoms. *Phys. Rev. A* **51**, 4704–4711 (1995).
- Bradley, C. C., Sackett, C. A. & Hulet, R. G. Bose-Einstein condensation of lithium: observation of limited condensate number. *Phys. Rev. Lett.* **78**, 985–989 (1997).
- Sackett, C. A., Gerton, J. M., Welling, M. & Hulet, R. G. Measurements of collective collapse in a Bose-Einstein condensate with attractive interactions. *Phys. Rev. Lett.* **82**, 876–879 (1999).
- Gerton, J. M., Strekalov, D., Prodan, I. & Hulet, R. G. Direct observation of growth and collapse of a Bose-Einstein condensate with attractive interactions. *Nature* **408**, 692–695 (2000).
- Donley, E. A. *et al.* Dynamics of collapsing and exploding Bose-Einstein condensates. *Nature* **412**, 295–299 (2001).
- Stegeman, G. I. & Segev, M. Optical spatial solitons and their interactions: Universality and diversity. *Science* **286**, 1518–1523 (1999).
- Dalfovo, F., Giorgini, S., Pitaevskii, L. P. & Stringari, S. Theory of Bose-Einstein condensation in trapped gases. *Rev. Mod. Phys.* **71**, 463–512 (1999).
- Burger, S., Bongs, K., Dettmer, S., Ertmer, W. & Sengstock, K. Dark solitons in Bose-Einstein condensates. *Phys. Rev. Lett.* **83**, 5198–5201 (1999).
- Denschlag, J. *et al.* Generating solitons by phase engineering of a Bose-Einstein condensate. *Science* **287**, 97–100 (2000).
- Anderson, B. P. *et al.* Watching dark solitons decay into vortex rings in a Bose-Einstein condensate. *Phys. Rev. Lett.* **86**, 2926–2929 (2001).
- Pérez-García, V., Michinel, H. & Herrero, H. Bose-Einstein solitons in highly asymmetric traps. *Phys. Rev. A* **57**, 3837–3842 (1998).
- Muryshv, A. E., van Linden van den Heuvell, H. B. & Shlyapnikov, G. V. Stability of standing matter waves in a trap. *Phys. Rev. A* **60**, R2665–R2668 (1999).
- Carr, L. D., Leung, M. A. & Reinhardt, W. P. Dynamics of the Bose-Einstein condensate: quasi-one-dimension and beyond. *J. Phys. B* **33**, 3983–4001 (2000).
- Kivshar, Y. S., Alexander, T. J. & Turitsyn, S. K. Nonlinear modes of a macroscopic quantum oscillator. *Phys. Lett. A* **278**, 225–230 (2001).
- Schreck, F. *et al.* Quasipure Bose-Einstein condensate immersed in a Fermi sea. *Phys. Rev. Lett.* **87**, 080403–1–080403–4 (2001).
- Görlitz, A. *et al.* Realization of Bose-Einstein condensates in lower dimensions. *Phys. Rev. Lett.* **87**, 130402–1–130402–4 (2001).
- Truscott, A. G., Strecker, K. E., McAlexander, W. I., Partridge, G. B. & Hulet, R. G. Observation of Fermi pressure in a gas of trapped atoms. *Science* **291**, 2570–2572 (2001).
- McAlexander, W. I. Collisional interactions in an Ultracold Lithium Gas. Thesis, Rice Univ. (2000).
- Tiesinga, E., Verhaar, B. J. & Stoof, H. T. C. Threshold and resonance phenomena in ultracold ground-state collisions. *Phys. Rev. A* **47**, 4114–4122 (1993).
- Inouye, S. *et al.* Observation of Feshbach resonances in a Bose-Einstein condensate. *Nature* **392**, 151–154 (1998).
- Roberts, J. L., Claussen, N. R., Cornish, S. L. & Wieman, C. E. Magnetic field dependence of ultracold collisions near a Feshbach resonance. *Phys. Rev. Lett.* **85**, 728–731 (2000).
- Zakharov, V. E. & Shabat, A. B. Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media. *Sov. Phys. JETP* **34**, 62–65 (1972).
- Carr, L. D., Clark, C. W. & Reinhardt, W. P. Stationary solutions of the one-dimensional nonlinear Schrödinger equation. II. Case of attractive nonlinearity. *Phys. Rev. A* **62**, 063611–1–063611–10 (2000).
- Tai, K., Hasegawa, A. & Tomita, A. Observation of modulational instability in optical fibers. *Phys. Rev. Lett.* **56**, 135–138 (1986).
- Gordon, J. P. Interaction forces among solitons in optical fibers. *Opt. Lett.* **8**, 596–598 (1983).
- Hasegawa, A. *Optical Solitons in Fibers* (Springer, New York, 1990).
- Lenz, G., Meystre, P. & Wright, E. M. Nonlinear atom optics. *Phys. Rev. Lett.* **71**, 3271–3274 (1993).
- Kasevich, M. A. Atom interferometry with Bose-Einstein condensed atoms. *C.R. Acad. Sci. IV* **2**, 497–507 (2001).

Acknowledgements

We thank W. I. McAlexander for providing the coupled channels calculation, B. Luey for making the magnetic coils, and T. Killian and H. Stoof for discussions. This work was supported by the US National Science Foundation, NASA, the Office of Naval Research and the Welch Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.G.H. (e-mail: randy@atomcool.rice.edu).

Spin-galvanic effect

S. D. Ganichev^{*†}, E. L. Ivchenko[†], V. V. Bel'kov[†], S. A. Tarasenko[†], M. Sollinger^{*}, D. Weiss^{*}, W. Wegscheider^{*‡} & W. Prettl^{*}

^{*} Fakultät für Physik, Universität Regensburg, D-93040 Regensburg, Germany
[†] A.F. Ioffe Physico-Technical Institute of the RAS, 194021 St Petersburg, Russia
[‡] Walter Schottky Institut, TU München, 85748 Garching, Germany

There is much recent interest in exploiting the spin of conduction electrons in semiconductor heterostructures together with their charge to realize new device concepts¹. Electrical currents are usually generated by electric or magnetic fields, or by gradients of, for example, carrier concentration or temperature. The electron spin in a spin-polarized electron gas can, in principle, also drive an electrical current, even at room temperature, if some general symmetry requirements are met. Here we demonstrate such a 'spin-galvanic' effect in semiconductor heterostructures, induced by a non-equilibrium, but uniform population of electron spins. The microscopic origin for this effect is that the two electronic sub-bands for spin-up and spin-down electrons are shifted in momentum space and, although the electron distribution in each sub-band is symmetric, there is an inherent asymmetry in the spin-flip scattering events between the two sub-bands. The resulting current flow has been detected by applying a magnetic field to rotate an optically oriented non-equilibrium spin polarization in the direction of the sample plane. In contrast to previous experiments, where spin-polarized currents were driven by electric fields in semiconductor^{2,3}, we have here the complementary situation where electron spins drive a current without the need of an external electric field.

Although it is usually ignored, it is well known that the spin degeneracy of sub-bands in semiconductor quantum well (QW) structures can be lifted, owing to terms linear in the wavevector $\mathbf{k}$ resulting from the spin orbit interaction in asymmetric potentials^{4,5}. For a two-dimensional electron gas (2DEG) system, this leads to the situation sketched in Fig. 1. The electron energy band splits into two sub-bands which are shifted in $\mathbf{k}$ -space and each of the bands comprise states with spin up or down. In this band structure spin polarization means that one sub-band is occupied up to higher energies than the other. This is depicted in Fig. 1, illustrating that there are more spin-down than spin-up electrons. Can such a situation of uniform spin polarization cause an electric current? As long as the carrier distribution in each sub-band is symmetric around the sub-band minimum at $k_{x\pm}$ no current flows. However, asymmetric $\mathbf{k}$ -dependent spin relaxation events can occur, indicated

by the dashed arrows in Fig. 1, which result in a current flow. This means that a current is driven by a homogeneous spin polarization. Below we describe the observation of such a current.

Phenomenologically, an electric current can be linked by general symmetry arguments to the electron's averaged spin S by

$$j_\alpha = \sum_\gamma Q_{\alpha\gamma} S_\gamma \quad (1)$$

where j is an electric current density, Q is a second-rank pseudo-tensor, and $\alpha, \gamma = 1, 2$ indicate coordinates. Non-vanishing tensor components $Q_{\alpha\gamma}$ can only exist in non-centrosymmetric systems belonging to one of the gyrotropic classes⁶. In zinc-blende-based heterojunctions with a 2DEG, non-zero components of $Q_{\alpha\gamma}$ exist in contrast to the corresponding bulk crystals⁷. In our (001)-grown heterojunctions with the C_{2v} symmetry only two linearly independent components, Q_{xy} and Q_{yx} , are different from zero ($x||[1\bar{1}0]$ and $y||[110]$) where x and y are cartesian coordinates. Hence, to observe a spin-polarization-driven current a spin component lying in the plane of the heterojunctions is required (for example, S_y in Fig. 2).

To achieve an in-plane spin orientation in experiment one could either use spin selective contacts⁸ (see Fig. 2a) or optical orientation⁹ by using circularly polarized light. Although significant progress concerning electrical spin injecting has been made recently^{10–12}, reliable spin-injection into lateral low-dimensional electron systems—at room temperature—is still a challenge. Furthermore, electrical spin injection causes, apart from the driving current, a laterally inhomogeneous spin polarization and hence additional driving forces for current flow which would hamper the unambiguous demonstration of the effect described here. Instead, we use optical spin orientation, which ensures a homogeneous non-equilibrium spin polarization and directly proves the spin-galvanic effect. In this case of exciting electrons from the valence band to the conduction band the conduction band gets selectively spin-populated owing to selection rules which allow transitions by circularly polarized light only if the spin of the electron is changed by ± 1 . In addition to this method we have also used circularly polarized terahertz radiation causing intraband instead of interband excitation. One interesting aspect of employing terahertz radiation is

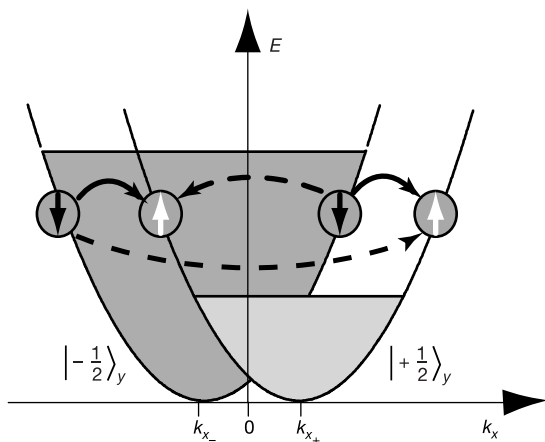


Figure 1 Microscopic origin of the spin-galvanic current in the presence of $\mathbf{k}$ -linear terms in the electron hamiltonian. The $\sigma_y k_x$ term in the hamiltonian splits the conduction band into two parabolas with the spin $\pm 1/2$ in the y direction. If one spin sub-band is preferentially occupied, for example, by spin injection (the $| -1/2 \rangle_y$ -states shown in the figure) asymmetric spin-flip scattering results in a current in the x direction. The rate of spin-flip scattering depends on the value of the initial and final $\mathbf{k}$ -vectors. There are four distinct spin-flip scattering events possible, indicated by the arrows. The transitions sketched by dashed arrows yield an asymmetric occupation of both sub-bands and hence a current flow. If, instead of the spin-down sub-band, the spin-up sub-band is preferentially occupied the current direction is reversed.

that only one type of carriers, electrons or holes, is involved. In this respect, the effect of intraband spin orientation¹³ is indeed very close to electrical spin injection.

It has been shown before that irradiation of quantum wells with circularly polarized light can result in a photocurrent caused by non-uniformly distributing photoexcited carriers in $\mathbf{k}$ -space according to optical selection rules and energy and momentum conservation. This is the circular photogalvanic effect observed recently in a 2DEG¹⁴. In order to achieve a uniform distribution in spin sub-bands and to exclude this circular photogalvanic effect we use the geometry depicted in Fig. 2b, where the photogalvanic current is identical to zero for normal incidence of the light¹⁴. In this geometry, optical excitation yields a steady-state spin orientation S_{0z} in the z direction with the generation rate $\dot{S}_z$ proportional to the intensity of the radiation. To obtain an in-plane component of the spins, necessary for the effect we describe here, a magnetic field, B , was applied (Fig. 2b). The field perpendicular to both the light-propagation direction $\hat{e}_z$ and the optically oriented spins rotates the spins into the plane of the 2DEG owing to Larmor precession. With the magnetic field oriented along the x axis we obtain a non-equilibrium spin polarization S_y which reads, after time averaging⁷:

$$S_y = - \frac{\omega_L \tau_{s\perp}}{1 + (\omega_L \tau_s)^2} S_{0z} \quad (2)$$

where $\tau_s = \sqrt{\tau_{s\parallel} \tau_{s\perp}}$ and $\tau_{s\parallel}, \tau_{s\perp}$ are the longitudinal and transverse electron-spin-relaxation times, $\omega_L = g\mu_B B_x / \hbar$ is the Larmor frequency, g is the in-plane effective electron g -factor, μ_B is the Bohr magneton, and $S_{0z} = \tau_{s\parallel} \dot{S}_z$ is the steady-state electron spin polarization in the absence of the magnetic field. Using the Larmor precession we prepared the situation sketched in Fig. 1 where the spin polarization S_y lies in the plane. The denominator in equation (2) yielding the decay of S_y for ω_L exceeding the inverse spin-relaxation time is well known from the Hanle effect¹⁵.

The experiments were carried out at room temperature ($T = 293$ K) and at liquid helium temperature on n -GaAs/AlGaAs single quantum wells of 15-nm width and on GaAs single heterojunctions. The (001)-oriented samples grown by molecular beam epitaxy contain 2DEG systems with electron densities $n_s \approx 2 \times 10^{11} \text{ cm}^{-2}$ and mobilities μ above $10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ at $T = 4.2$ K. Two pairs of contacts were centred on opposite sample edges along the directions $x||[1\bar{1}0]$ and $y||[110]$ (see inset in Fig. 3). Complementary measurements were also carried out on p -GaAs multiple quantum-well structures containing 20 wells of 15-nm width with hole densities in each well $p_s = 2 \times 10^{11} \text{ cm}^{-2}$ and $\mu = 5 \times 10^5 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$.

At room temperature a magnetic field of up to 1 T was generated by an electromagnet. For the low-temperature measurements the samples were placed in a cryostat with a split-coil superconducting magnet yielding a field B of up to 3 T. For optical interband excitation a continuous-wave Ti:sapphire laser was used at a wavelength of $\lambda = 0.777 \mu\text{m}$. In order to extract the spin-galvanic current the linearly polarized laser beam was transmitted through a

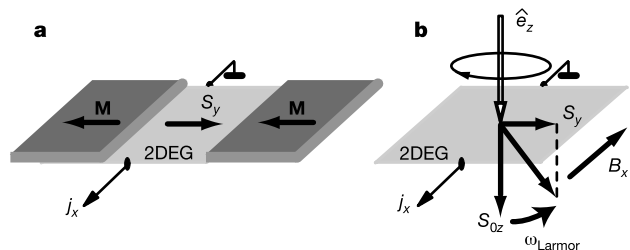


Figure 2 Two ways of generating an in-plane spin-polarization. **a**, The spin injection from ferromagnetic contacts of magnetization $\mathbf{M}$ into the two-dimensional electron gas (2DEG). **b**, The optical orientation in combination with an in-plane magnetic field B_x . Spins, initially aligned along the z direction are rotated into the y direction by B_x .

photoelastic modulator which yields a periodically oscillating polarization between right-circular and left-circular. The current was recorded by a lock-in amplifier in phase with the modulator. For intraband excitation we used the $\lambda = 148 \mu\text{m}$ line of a pulsed ammonia laser¹⁶. The current of unbiased samples was measured by the voltage drop on a 50Ω load resistor in a closed circuit configuration.

Both for visible and terahertz radiation a current has been observed for all *n*-type samples after applying an in-plane magnetic field (Figs 3 and 4). For small magnetic fields B where $\omega_L \tau_s < 1$ holds, the spin-galvanic current increases linearly as expected from equations (1) and (2). This is seen in the room-temperature data of Fig. 3 and in the 4.2 K data in Fig. 4 for $B \leq 1$ T. The polarity of the current depends on the direction of the excited spins ($\pm z$ -direction for right- or left-circularly polarized light, respectively) and on the direction of the applied magnetic field ($\pm x$ direction). The current is parallel (or anti-parallel) to the magnetic field vector. Comparing the power sensitivity for visible and terahertz excitation we find them to be of the same order of magnitude as seen in Fig. 3. However, we note that the current contribution per photon is larger by two orders of magnitude for interband excitation. This is due to a more effective spin generation rate by interband excitation. It may even be larger, because the signal gets partially shortened by photogenerated carriers in the semi-insulating substrate.

For higher magnetic fields the current saturates and decreases upon further increase of B , as shown in Fig. 4. This is ascribed to the Hanle effect, see equation (2). The observation of the Hanle effect demonstrates that free carrier intraband transitions can polarize the spins of electron systems. The measurements allow us to obtain the spin-relaxation time τ_s from the peak position of the photocurrent where $\omega_L \tau_s = 1$ holds if the *g*-factor is known. The *g*-factor of quantum wells, however, depends strongly on structural details. This leads to an uncertainty in the measured spin lifetime itself. Using the theoretically estimated¹⁷ value of $g = -0.2$ results in $\tau_s = 40$ ps at 4.2 K. Spin-relaxation times of the same order of magnitude as derived here have been obtained by photoluminescence measurements¹⁸. However, direct comparison of these results with our data is difficult, because optical recombination does not detect the spin-relaxation time at the Fermi energy of the 2DEG, as in the present case of monopolar spin orientation. Published data

on Faraday rotation experiments, which give access to spin relaxation at the Fermi level, however, refer to materials which are not directly comparable to our samples^{19,20}.

Because the in-plane *g*-factor for heavy holes is very small²¹, the associated spin-galvanic effect is expected to be negligible. Indeed no magnetic-field-induced current has been observed in *p*-GaAs quantum wells at terahertz excitation causing spin polarization of holes only. We note that a magnetic-field-induced circular photogalvanic effect in *p*-type bulk materials was previously observed in GaAs at intraband excitation²². However, this effect is not due to spin orientation and does not occur in *p*-type quantum wells owing to spatial quantization²³. Only for visible excitation, which polarizes both electrons and holes, has a current signal been detected in *p*-type samples (see Fig. 3). This current is due to the spin polarization of electrons again, which are in this case the minority carriers generated by interband excitation.

Microscopically, the spin-galvanic effect is caused by the asymmetric spin-flip relaxation of spin-polarized electrons in systems with *k*-linear contributions to the effective hamiltonian. The lifting of spin degeneracy of 2DEG depicted in Fig. 1 is a consequence of a contribution to the hamiltonian of the form $\hat{H}_k = \sum_{\alpha\gamma} \beta_{\alpha\gamma} \sigma_{\alpha} k_{\gamma}$ where σ_{α} are the Pauli spin matrices and $\beta_{\alpha\gamma}$ is a pseudo-tensor subjected to the same symmetry restriction as the pseudo-tensor $Q_{\alpha\gamma}$ used in equation (1). Figure 1 sketches the electron energy spectrum with the $\beta_{yx} \sigma_y k_x$ term included. Spin orientation in the *y* direction causes an unbalanced population in the spin-down and spin-up sub-bands. The current flow is caused by the *k*-dependent spin-flip relaxation processes. Spins oriented in the *y* direction are scattered along k_x from the higher filled (for example, spin-down) sub-band, $| -1/2 \rangle_y$, to the less filled spin-up sub-band, $| +1/2 \rangle_y$. Four quantitatively different spin-flip scattering events exist and are sketched in Fig. 1 by bent arrows. The spin-flip scattering rate depends on the values of the wavevectors of the initial and the final states, respectively²⁴. Therefore spin-flip transitions, shown by solid arrows in Fig. 4, have the same rates. They preserve the symmetric distribution of carriers in the sub-bands and, thus, do not yield a current. However, the two scattering processes shown by broken arrows are not equivalent and generate an asymmetric carrier distribution around the sub-band minima in both sub-bands. This asymmetric distribution results in a current flow along the *x* direction. Within our model of elastic scattering the current is not

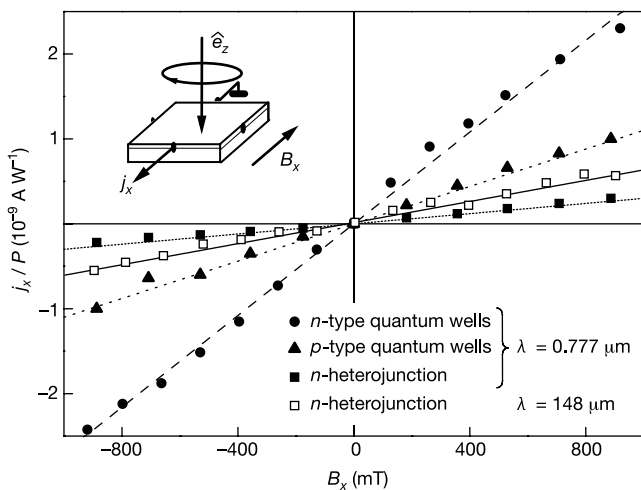


Figure 3 Current density j_x normalized to the radiation power P as a function of the magnetic field B for normally incident circularly polarized radiation at room temperature for various samples and wavelengths. Filled symbols: $\lambda = 0.777 \mu\text{m}$, $P = 100 \text{ mW}$. Triangles, squares and circles correspond to *n*-type and *p*-type multiple quantum wells, and an *n*-type GaAs/AlGaAs heterojunction, respectively. Open squares: *n*-type GaAs/AlGaAs heterojunction, $\lambda = 148 \mu\text{m}$, $P = 20 \text{ kW}$. The inset shows the geometry of the experiment where $\hat{e}_z$ indicates the direction of the incoming light.

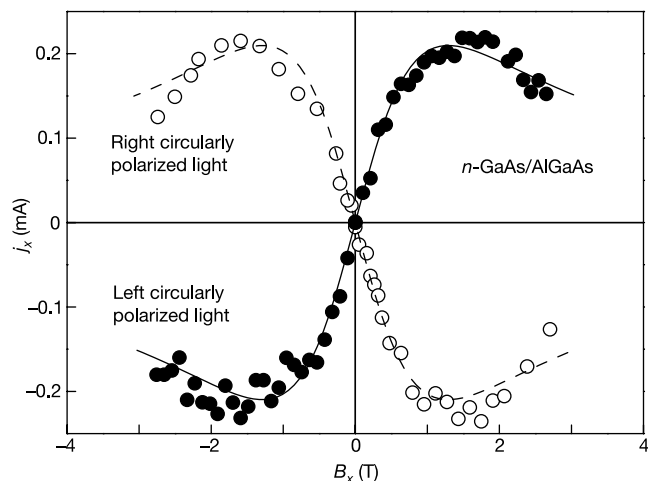


Figure 4 Current j_x as a function of magnetic field B for normally incident right-handed (open circles) and left-handed (filled circles) circularly polarized radiation at $\lambda = 148 \mu\text{m}$ and radiation power 20 kW. Measurements are presented for a *n*-GaAs/AlGaAs single heterojunction at $T = 4.2 \text{ K}$. Curves are fitted from equation (2) using the same value of the spin relaxation time τ_s and scaling of the j_x value for the solid curve as for the dashed curve.

spin-polarized because the same number of spin-up and spin-down electrons move in the same direction with the same velocity. In the case of inelastic scattering the current may be polarized.

The uniformity of the spin polarization in space is preserved during the scattering processes. Therefore, the spin-galvanic effect differs from surface currents induced by inhomogeneous spin orientation²⁵. It also differs from other experiments where the spin current is caused by gradients of potentials, concentrations and so on, like the spin-voltaic effect^{26,27} and the photo-voltaic effect, which occur in inhomogeneous samples, and like the 'paramagnetic metal-ferromagnet' junction or *p-n* junctions. When considering spintronic devices involving heterojunctions, spin-galvanic current must be taken into account. □

Received 3 January; accepted 25 March 2002.

1. Wolf, S. A. *et al.* Spintronics: a spin-based electronics vision for the future. *Science* **294**, 1488–1495 (2001).
2. Hägele, D. *et al.* Spin transport in GaAs. *Appl. Phys. Lett.* **73**, 1580–1582 (1998).
3. Kikkawa, J. M. & Awschalom, D. D. Lateral drag of spin coherence in gallium arsenide. *Nature* **397**, 139–141 (1999).
4. Bychkov, Y. A. & Rashba, E. I. Properties of a 2D electron gas with lifted spectral degeneracy. *Sov. JETP Lett.* **39**, 78–81 (1984).
5. D'yakov, M. I. & Kocharovskii, V. Yu. Spin relaxation of two-dimensional electrons in noncentrosymmetric semiconductors. *Sov. Phys. Semicond.* **20**, 110–111 (1986).
6. Koopmans, B., Santos, P. V. & Cardona, M. Optical activity in semiconductors: stress and confinement effects. *Phys. Status Solidi* **205**, 419–463 (1998).
7. Ivchenko, E. L., Lyanda-Geller, Yu. B. & Pikus, G. E. Current of thermalized spin-oriented photocarriers. *Sov. Phys. JETP* **71**, 550–557 (1990).
8. Fiederling, R. *et al.* Injection and detection of spin-polarized current in a light-emitting diode. *Nature* **402**, 787–789 (1999).
9. Meier, F. & Zakharichenya, B. P. (eds) *Optical Orientation* 1–523 (Elsevier Science, Amsterdam, 1984).
10. Hammar, P. R. & Johnson, M. Spin-dependent current transmission across a ferromagnet-insulator-two-dimensional electron gas junction. *Appl. Phys. Lett.* **79**, 2591–2593 (2001).
11. Zhu, H. J. *et al.* Room-temperature spin injection from Fe into GaAs. *Phys. Rev. Lett.* **87**, 016601–016601-4 (2001).
12. Hanbicki, A. T. *et al.* Efficient electrical spin injection from a magnetic metal/tunnel barrier contact into a semiconductor. *Appl. Phys. Lett.* **80**, 1240–1242 (2002).
13. Tarasenko, S. A. & Ivchenko, E. L. Spin orientation of two-dimensional electron gas under intraband optical pumping. Preprint cond-mat/0202471 at (<http://xxx.lanl.gov>) (2002).
14. Ganichev, S. D. *et al.* Conversion of spin into direct electric current in quantum wells. *Phys. Rev. Lett.* **86**, 4358–4361 (2001).
15. Hanle, W. Über magnetische Beeinflussung der Polarisation der Resonanzfluoreszenz. *Z. Phys.* **30**, 93–105 (1924).
16. Ganichev, S. D. Tunnel ionization of deep impurities in semiconductors induced by terahertz electric fields. *Physica B* **273–274**, 737–742 (1999).
17. Ivchenko, E. L., Kiselev, A. A. & Willander, M. Electronic g-factor in biased quantum wells. *Solid State Commun.* **102**, 375–378 (1997).
18. Damen, T. C. *et al.* Subpicosecond spin relaxation dynamics of excitons and free carriers in GaAs quantum wells. *Phys. Rev. Lett.* **67**, 3432–3435 (1991).
19. Kikkawa, J. M. *et al.* Room-temperature spin memory in two-dimensional electron gases. *Science* **277**, 1284–1287 (1997).
20. Sandu, J. S. *et al.* Gateable suppression of spin relaxation in semiconductors. *Phys. Rev. Lett.* **86**, 2150–2153 (2001).
21. Marie, X. *et al.* Hole spin quantum beats in quantum-well structures. *Phys. Rev. Lett.* **60**, 5811–5817 (1999).
22. Andrianov, A. V. & Yaroshetskii, I. D. Magnetic-field-induced circular photovoltaic effect in semiconductors. *Sov. JETP Lett.* **40**, 882–884 (1984).
23. Ivchenko, E. L., Lyanda-Geller, Yu. B. & Pikus, G. E. Circular magnetophotocurrent and spin splitting of band states in optically-inactive crystals. *Solid State Commun.* **69**, 663–665 (1989).
24. Averkiev, N. S., Golub, L. E. & Willander, M. Spin relaxation anisotropy in two-dimensional semiconductor systems. Preprint cond-mat/0202437 at (<http://xxx.lanl.gov>) (2002).
25. Averkiev, N. S. & D'yakov, M. I. Current due to inhomogeneity of the spin orientation of electrons in a semiconductor. *Sov. Phys. Semicond.* **17**, 393–395 (1983).
26. Johnson, M. & Silsbee, R. H. Interfacial charge-spin coupling: injection and detection of spin magnetization in metals. *Phys. Rev. Lett.* **55**, 1790–1793 (1985).
27. Zutic, I., Fabian, J. & Das Sarma, S. Spin-polarized transport in inhomogeneous magnetic semiconductors: theory of magnetic/nonmagnetic *p-n* junctions. *Phys. Rev. Lett.* **88**, 066603–066603-4 (2001).

Acknowledgements

We thank D. I. Kovalev, W. Schoepe and M. Bichler for helpful discussions and support. We acknowledge financial support from the DFG, the RFFI and INTAS.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.D.G. (e-mail: sergey.ganichev@physik.uni-regensburg.de).

Terahertz semiconductor-heterostructure laser

Rüdiger Köhler*, Alessandro Tredicucci*, Fabio Beltram*, Harvey E. Beere†, Edmund H. Linfield†, A. Giles Davies†, David A. Ritchie†, Rita C. Iotti‡ & Fausto Rossi‡

* NEST-INFM and Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy

† Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK

‡ INFM and Dipartimento di Fisica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

Semiconductor devices have become indispensable for generating electromagnetic radiation in everyday applications. Visible and infrared diode lasers are at the core of information technology, and at the other end of the spectrum, microwave and radio-frequency emitters enable wireless communications. But the terahertz region (1–10 THz; 1 THz = 10^{12} Hz) between these ranges has remained largely underdeveloped, despite the identification of various possible applications—for example, chemical detection, astronomy and medical imaging^{1–4}. Progress in this area has been hampered by the lack of compact, low-consumption, solid-state terahertz sources^{5–9}. Here we report a monolithic terahertz injection laser that is based on interminiband transitions in the conduction band of a semiconductor (GaAs/AlGaAs) heterostructure. The prototype demonstrated emits a single mode at 4.4 THz, and already shows high output powers of more than 2 mW with low threshold current densities of about a few hundred A cm^{−2} up to 50 K. These results are very promising for extending the present laser concept to continuous-wave and high-temperature operation, which would lead to implementation in practical photonic systems.

In conventional semiconductor lasers, light is generated by the radiative recombination of conduction band electrons with valence band holes across the bandgap of the active material; in contrast, electrons in a quantum-cascade laser propagate through a potential staircase of coupled quantum wells, where the conduction band is split by quantum confinement into a number of distinct sub-bands¹⁰. By choice of layer thickness and applied electric field, lifetimes and tunnelling probabilities of each level are engineered in order to obtain population inversion between two sub-bands in a series of identical repeat units. Injector/collector structures connect these active regions, allowing electrical transport through injection of carriers into the upper laser level, and extraction of carriers from the lower laser level. The radiation frequency is determined by the energy spacing of the lasing sub-bands, allowing in principle operation at arbitrarily long wavelengths. The quantum-cascade scheme has thus long been the preferred choice in many attempts to fabricate a terahertz semiconductor laser. Although electroluminescent devices have been reported by several groups^{11–14}, laser action has been shown only at much shorter wavelengths^{15,16}. In fact, above the forbidden phonon band of the material, direct electron–longitudinal optical (LO) phonon scattering processes can be conveniently used to achieve large population inversions¹⁶. Furthermore, an additional problematic issue for the terahertz range stems from the fact that conventional laser waveguides are not suitable, owing to large free-carrier absorption losses and practical limitations on the thickness of epilayer growth.

As in all lasers, efficient depletion of the lower level is essential, and long lifetimes of the upper level are highly desirable. Up until now, proposed terahertz quantum-cascade designs focused mainly on the latter aspect. To this end, structures have featured narrow injector minibands to both suppress scattering of electrons from the

upper laser state through LO-phonon emission and block cross-absorption of the emitted light. This, however, results in relatively small tunnel coupling between the active region sub-bands and the injector states. As a consequence, extraction of carriers from the lower laser level is very slow, which, compounded with the limited current densities supported by the narrow injector minibands, hinders the achievement of population inversion¹⁷. On the contrary, our design targets precisely these two issues, an objective best fulfilled by the adoption of chirped superlattice active regions¹⁸. In particular, one repeat unit in our structure comprises seven GaAs quantum wells separated by $\text{Al}_{0.15}\text{Ga}_{0.85}\text{As}$ barriers, with the active region consisting of three closely coupled quantum wells.

A self-consistent calculation of the wavefunctions and energies in our device is shown in Fig. 1a, and details of the design are given in

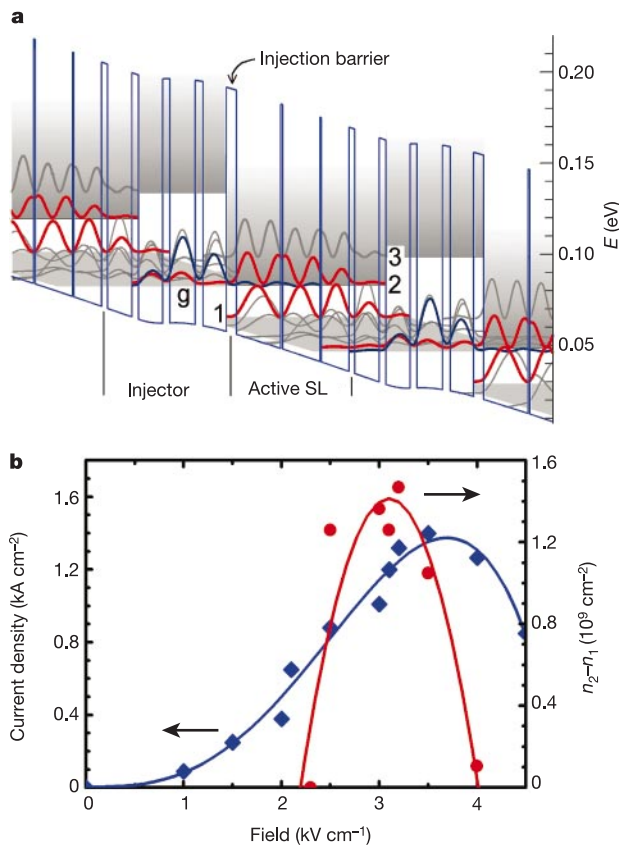


Figure 1 Conduction band structure of the laser active core and computed population inversion and current density at several applied biases. **a**, Self-consistent calculation of the conduction band structure of a portion of the layer stack in the waveguide core under a field of 3.5 kV cm^{-1} . Injectors and superlattice (SL) active regions are alternating. The layer thicknesses, in nm, starting from the injection barrier are **4.3/ 18.8/ 0.8/ 15.8/ 0.6/ 11.7/ 2.5/ 10.3/ 2.9/ 10.2/ 3.0/ 10.8/ 3.3/ 9.9**, where $\text{Al}_{0.15}\text{Ga}_{0.85}\text{As}$ layers are in bold face, the active region is in italic, and the 10.2-nm-wide well is Si-doped at $4 \times 10^{16} \text{ cm}^{-3}$. The moduli squared of the wavefunctions are shown, with miniband regions represented by shaded areas. The optical transition takes place across the 18-meV-wide minigap between the second and first miniband (states 2 and 1) and, being vertical in real space, presents a large dipole matrix element of 7.8 nm. Carriers are injected into state 2 via resonant tunnelling from the injector ground state labelled g. **b**, Calculated population inversion between states 2 and 1 of the active superlattice (circles), expressed in terms of electron sheet density (n_2 and n_1 respectively), and current density traversing the structure (diamonds). Both are computed as a function of bias field at 10 K. Solid lines are polynomial fits. From the simulation we also extract the level lifetimes: $\tau_2 = 0.8 \text{ ps}$, $\tau_{21} = 8.3 \text{ ps}$, $\tau_1 = 2.2 \text{ ps}$ with τ_2 and τ_{21} dominated by carrier-phonon scattering processes. Note that τ_{21} is much greater than τ_2 owing to the superlattice nature of the active region. This implies that $\tau_1 < \tau_{21}$, which is the necessary condition for population inversion.

Fig. 1 legend. The lower laser state 1 is strongly coupled to a wide injector miniband, comprising seven sub-bands spanning an energy of 17 meV. This provides a large phase space where electrons scattered either from sub-band 2 or directly from the injector can spread, at the same time ensuring fast depletion of state 1. Moreover, the wide miniband allows efficient electrical transport, even at high current densities, and simultaneously suppresses thermal backfilling. The validity of this design is supported by theoretical modelling¹⁷ employing a Monte Carlo scheme based on a coupled set of fully three-dimensional Boltzmann equations¹⁹, including all relevant energy-relaxation mechanisms, like carrier-carrier and carrier-LO-phonon scattering processes²⁰. The use of suitable periodic boundary conditions allows simulation of sub-band populations as well as current voltage characteristics without resorting to phenomenological parameters. The results displayed in Fig. 1b indeed predict the build-up of a significant population inversion above an applied electric field of 2 kV cm^{-1} , peaking at about $1.5 \times 10^9 \text{ cm}^{-2}$ just before the design field of 3.5 kV cm^{-1} . Large current densities of the order of 1 kA cm^{-2} are also obtained before the onset of negative differential resistance that marks the end of resonant tunnelling transport. It is worth noting that, contrary to mid-infrared quantum-cascade lasers, where depletion of the lower laser state is controlled by optical phonon emission, in this case the

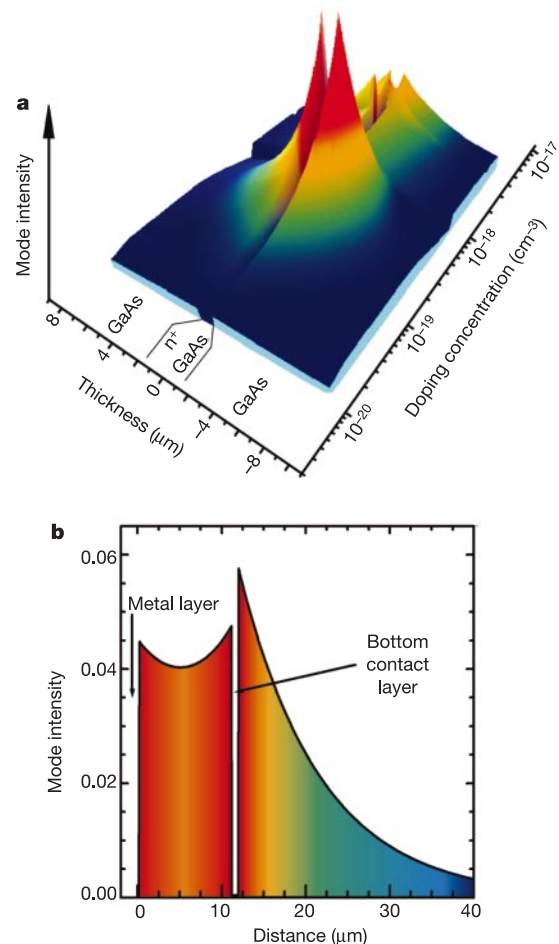


Figure 2 Waveguide design principle and radiation mode confinement. **a**, Evolution of the intensity profile of the optical mode bound to a 800-nm-thick GaAs n-doped layer within undoped GaAs material calculated as a function of donor density. The dielectric constants were derived from a Drude-Lorentz model. **b**, Calculated mode profile along the growth direction of the final device structure. The origin of the abscissa is at the top metal-semiconductor interface, and the waveguide core of 104 repetitions of the active region shown in Fig. 1 is between the bottom contact and top metal layer.

transparency condition of the interminiband transition is reached at a current density significantly different from zero. Moreover, in our calculations population inversion is reached only in the presence of carrier–carrier scattering. This process provides the necessary in-plane momentum to activate relaxation of electrons in state 1 via carrier–phonon scattering¹⁷.

As stated above, the waveguide is a very important aspect of the present device. Traditionally, long-wavelength quantum-cascade lasers exploit the surface plasmon mode at the interface between two materials with dielectric constants of opposite sign (normally a metal and a semiconductor) to achieve a tight optical confinement with low absorption losses^{15,21}. At terahertz frequencies, however, a simple surface-plasmon waveguide results in a small overlap of the optical mode with any active region of reasonable thickness. Therefore, more sophisticated solutions have been considered, including double-surface plasmon structures where the guided mode is totally contained between two metallic claddings²². Here we have chosen to use a thin (800 nm) n^+ GaAs layer between 104 repeat periods of the above active superlattice on one side and an undoped GaAs substrate on the other in order to create a strongly confined low loss TM mode bound to that layer. This is a result of the n^+ layer dielectric constant (ϵ_A), which, by appropriately tuning the doping density, can be made negative, albeit comparable in modulus to that of the surrounding semiconductor (ϵ_B). The two surface plasmons existing at the two layer interfaces merge into a single mode because of the relatively small thickness of the n^+ layer with respect to the

surface plasmon skin depth (δ_A):

$$\delta_A = \frac{\lambda}{2} \left| \operatorname{Re} \left[\epsilon_A \sqrt{\frac{-1}{\epsilon_A + \epsilon_B}} \right] \right|^{-1}$$

The penetration of this mode into the outside semiconductor is at the same time minimized, being to a first approximation proportional to $\sqrt{-(\epsilon_A + \epsilon_B)}$, resulting in a very tight confinement.

In fact the situation is more intricate, owing to the complex nature of the dielectric constants, as can be seen in Fig. 2a where the evolution of the mode profile is plotted as a function of the n^+ layer doping density. We chose a donor density $n = 2 \times 10^{18} \text{ cm}^{-3}$, which yields a good compromise between absorption losses and overlap with the active material. Additionally this makes it possible to use the n^+ layer for electrical contacting. The presence on top of the active regions of a 200-nm-thick GaAs layer doped to $n = 5 \times 10^{18} \text{ cm}^{-3}$ with its metallic contact leads to the final mode profile of the actual device shown in Fig. 2b. A confinement factor $\Gamma = 0.47$ for the active material and an attenuation coefficient $\alpha_w = 16 \text{ cm}^{-1}$ are computed, which are extremely favourable values at these long wavelengths.

The sample was grown by molecular beam epitaxy and further processed into mesa-etched ridge stripes by optical contact lithography and wet chemical etching. It was then thinned down to about 200 μm in order to improve heat dissipation, and top and bottom Ge/Au ohmic contacts were provided by thermal evaporation and subsequent annealing under a nitrogen atmosphere. A Fabry–Perot cavity was formed by cleaving the stripes into laser bars with the facets left uncoated. The devices were finally soldered, using an In–Ag alloy, onto a copper block, wire-bonded, and mounted onto the cold finger of a liquid-helium flow cryostat equipped with polyethylene windows.

Figure 3 shows the emission spectra of one such laser at 8 K for different drive currents. The characteristic narrowing of the emis-

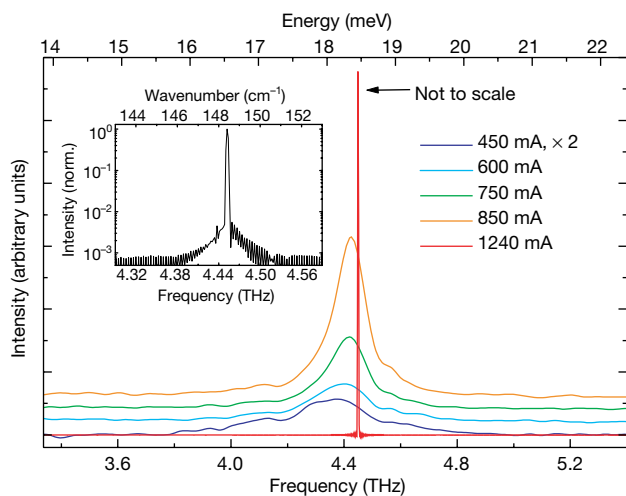


Figure 3 Emission spectra from a 1.24-mm-long and 180- μm -wide laser device recorded at 8 K for different drive currents. The lowest curve is multiplied by a factor of two for clarity, while the 1,240-mA laser spectrum (shown red) is scaled down by several orders of magnitude. Current macro-pulses were applied at a repetition rate of 333 Hz and matched the frequency response of the detector; each macro-pulse comprised 750 micro-pulses of 200 ns width at intervals of 2 μs to avoid excess heating in the device. The emitted radiation was collected by an $f/1$ off-axis parabolic mirror, passed through a Fourier-transform interferometer (FTIR) and focused again onto a liquid-helium-cooled Si bolometer. The entire beam path was purged with purified air in order to minimize water-vapour absorption. Subthreshold spectra were recorded with a resolution of 2 cm^{-1} (60 GHz) using the FTIR in step-scan mode and employing a lock-in amplifier for signal detection. The spontaneous emission at low drive currents (300 mA) displays a full-width at half-maximum of 2 meV, similar to the value obtained from previous luminescence structures with only 40 periods²³. It is best fitted with a lorentzian line shape, which, together with the absence of any broadening when increasing the number of superlattice periods, indicates the high quality and uniformity of the growth throughout the whole 10- μm -thick active material stack. Laser spectra were collected in rapid-scan mode with the maximum possible resolution of 0.125 cm^{-1} (3.75 GHz) using a DTGS (deuterated triglycine sulphate) detector. Inset, the laser line on a logarithmic scale showing single-mode emission with a side-mode suppression ratio of 20 dB.

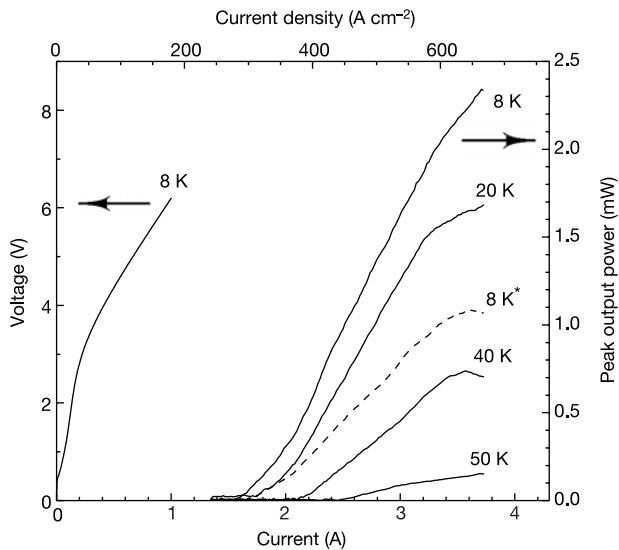


Figure 4 Light–current (L – I) characteristics of a 180- μm -wide and 3.1-mm-long laser ridge. Data were recorded in pulsed mode, applying 100-ns-long pulses at a repetition rate of 333 Hz. The peak power values represent what was collected from a single facet onto our detector after correction by a factor of 2 to account for the FTIR retro-reflection. At the lowest temperature the device emits more than 2 mW. When the duty cycle was increased to 0.5%, the peak power reduced to 1 mW owing to the stronger heating of the device, as shown by the dashed curve labelled with an asterisk. At the maximum operating temperature of 50 K, we observed 120 μW of peak output power. The optical output power was estimated using the responsivity of the bolometer as calibrated against black-body radiation by the manufacturer (QMC Instruments). The V – I curve at a temperature of 8 K is also displayed.

sion line and nonlinear dependence of the intensity are clearly observed up to a current of about 880 mA, where laser threshold is reached. Lasing takes place at ~ 4.4 THz, on the high-energy side of the luminescence line, probably owing to the reduced waveguide losses at shorter wavelengths. Single-mode emission is obtained, which is probably a consequence of the relatively narrow gain spectrum and the wide Fabry–Perot mode spacing. A high-resolution laser spectrum is shown in the inset; the measured linewidth is limited by the resolution of the spectrometer (3.75 GHz).

Figure 4 shows the light–current (L – I) and voltage–current (V – I) characteristics of a representative device. At a heat-sink temperature of 8 K, the output peak power is estimated to be more than 2 mW, with a threshold current density of 290 A cm^{-2} . The latter is a very small value for quantum-cascade lasers and allows operation at high duty cycles (up to 10%) even in this large device. We expect that narrower stripes and appropriate changes in sample processing would readily lead to continuous-wave operation. The initial high resistivity in the V – I characteristics stems from misalignment of the sub-bands at low field; at higher fields, the injector miniband lines up with the second miniband of the following stage and carrier injection into the upper laser level takes place. The electric field at threshold is 7.5 kV cm^{-1} . This is larger than the design value of 3.5 kV cm^{-1} , probably as a result of the non-negligible contact resistance. As expected, a negative differential resistance feature was observed at about 850 A cm^{-2} .

These experimental results match well the theoretical predictions of Fig. 1b. The V – I curve has all the distinctive qualitative features and the measured current densities are of the same order of the computed ones, showing that even at these small energies carrier relaxation and transport are dominated by electron–LO phonon and electron–electron scattering. The discrepancy is a factor of about 1.5, possibly related to acoustic phonon or impurity scattering or to a lower than specified free-carrier density in the sample. In fact, the simulation indicates that a reduction in the doping density of the injectors by 25% would lead to a reduction in current density of 35%. From the theoretical values of population inversion and confinement factor, we calculate a maximum modal gain of 23 cm^{-1} (ref. 18). This value compares well with the estimated cavity losses $\alpha_W + \alpha_M = (16 + 4) \text{ cm}^{-1} = 20 \text{ cm}^{-1}$, α_M being the mirror out-coupling of the 3.1-mm-long stripe. This consistency is confirmed by the experimental observation that laser stripes shorter than 1 mm, with corresponding larger mirror losses, do not reach laser threshold.

We believe that improved design of the active region of our device (in particular aiming at the reduction of thermal backfilling), together with optimized fabrication (junction-down mounting, facet coating, lateral overgrowth), would rapidly lead to continuous-wave emission and to operation at liquid-nitrogen temperatures. The present demonstration of a terahertz quantum-cascade laser, operating below the LO phonon band, represents a first step towards the development of widely usable terahertz photonics. □

Received 14 January; accepted 19 February 2002.

1. Miles, R. E., Harrison, P. & Lippens, D. (eds) *Terahertz Sources and Systems* Vol. 27 (NATO Science Series II, Kluwer, Dordrecht, 2001).
2. Han, P. Y., Cho, G. C. & Zhang, X.-C. Time-domain transillumination of biological tissue with terahertz pulses. *Opt. Lett.* **25**, 242–244 (2000).
3. Mittleman, D. M., Jacobsen, R. H. & Nuss, M. C. T-ray imaging. *IEEE J. Sel. Top. Quant. Electron.* **2**, 679–692 (1996).
4. Mittleman, D. M., Hunsche, S., Boivin, L. & Nuss, M. C. T-ray tomography. *Opt. Lett.* **22**, 904–906 (1997).
5. Pavlov, S. G. *et al.* Stimulated emission from donor transitions in silicon. *Phys. Rev. Lett.* **84**, 5220–5223 (2000).
6. Matsuura, S., Tani, M. & Sakai, K. Generation of coherent terahertz radiation by photomixing in dipole photoconductive antennas. *Appl. Phys. Lett.* **70**, 559–561 (1997).
7. Hu, B. B., Zhang, X.-C. & Auston, D. H. Terahertz radiation induced by subband-gap femtosecond optical excitation of GaAs. *Phys. Rev. Lett.* **67**, 2709–2712 (1991).
8. Kersting, R., Unterrainer, K., Strasser, G., Kauffmann, H. F. & Gornik, E. Few-cycle THz emission from cold plasma oscillations. *Phys. Rev. Lett.* **79**, 3038–3041 (1997).
9. Bründermann, E., Chamberlin, D. R. & Haller, E. E. High duty cycle and continuous terahertz emission from germanium. *Appl. Phys. Lett.* **76**, 2991–2993 (2000).

10. Faist, J. *et al.* Quantum cascade laser. *Science* **264**, 553–556 (1994).
11. Rochat, M., Faist, J., Beck, M., Oesterle, U. & Illegems, M. Far-infrared ($\lambda = 88 \mu\text{m}$) electroluminescence in a quantum cascade structure. *Appl. Phys. Lett.* **73**, 3724–3726 (1998).
12. Williams, B. S., Xu, B., Hu, Q. & Melloch, M. R. Narrow-linewidth terahertz intersubband emission from three-level systems. *Appl. Phys. Lett.* **75**, 2927–2929 (1999).
13. Ulrich, J., Zobl, R., Schrenk, W., Strasser, G. & Unterrainer, K. Terahertz quantum cascade structures: Intra- versus interwell transition. *Appl. Phys. Lett.* **76**, 1928–1930 (2000).
14. Helm, M., England, P., Colas, E., De Rosa, F. & Allen, S. J. Intersubband emission from semiconductor superlattices excited by sequential resonant tunnelling. *Phys. Rev. Lett.* **63**, 74–77 (1989).
15. Colombelli, R. *et al.* Far-infrared surface-plasmon quantum-cascade lasers at $21.5 \mu\text{m}$ and $24 \mu\text{m}$ wavelengths. *Appl. Phys. Lett.* **78**, 2620–2622 (2001).
16. Hofstetter, D., Beck, M., Aellen, T. & Faist, J. High-temperature operation of distributed feedback quantum-cascade lasers at $5.3 \mu\text{m}$. *Appl. Phys. Lett.* **78**, 396–398 (2001).
17. Köhler, R., Iotti, R. C., Tredicucci, A. & Rossi, F. Design and simulation of terahertz quantum cascade lasers. *Appl. Phys. Lett.* **79**, 3920–3922 (2001).
18. Tredicucci, A. *et al.* High performance interminiband quantum cascade lasers with graded superlattices. *Appl. Phys. Lett.* **73**, 2101–2103 (1998).
19. Iotti, R. C. & Rossi, F. Nature of charge transport in quantum-cascade lasers. *Phys. Rev. Lett.* **87**, 146603–1–146603–4 (2001).
20. Iotti, R. C. & Rossi, F. Carrier thermalization versus phonon-assisted relaxation in quantum-cascade lasers: A Monte Carlo approach. *Appl. Phys. Lett.* **78**, 2902–2904 (2001).
21. Sirtori, C. *et al.* Long-wavelength ($\lambda \sim 8$ – $11.5 \mu\text{m}$) semiconductor lasers with waveguides based on surface plasmons. *Opt. Lett.* **23**, 1366–1368 (1998).
22. Rochat, M., Beck, M., Faist, J. & Oesterle, U. Measurement of far-infrared waveguide loss using a multisection single-pass technique. *Appl. Phys. Lett.* **78**, 1967–1969 (2001).
23. Köhler, R. *et al.* High-intensity interminiband terahertz emission from chirped superlattices. *Appl. Phys. Lett.* **80**, 1867 (2002).

Acknowledgements

We thank S. Dhillon for discussions. This work was supported in part by the European Commission through the IST Framework V FET project WANTED. R.K. was supported by the C.N.R.; E.H.L. and A.G.D. were supported by Toshiba Research Europe Ltd and The Royal Society, respectively.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.K. (e-mail: koehler@nest.sns.it).

Ocean productivity before about 1.9 Gyr ago limited by phosphorus adsorption onto iron oxides

Christian J. Bjerrum*†‡ & Donald E. Canfield*

* Danish Center for Earth System Science, Institute of Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense, Denmark

† Danish Center for Earth System Science, Niels Bohr Institute for Astronomy, Physics and Geophysics, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark

After the evolution of oxygen-producing cyanobacteria at some time before 2.7 billion years ago¹, oxygen production on Earth is thought to have depended on the availability of nutrients in the oceans, such as phosphorus (in the form of orthophosphate). In the modern oceans, a significant removal pathway for phosphorus occurs by way of its adsorption onto iron oxide deposits^{2,3}. Such deposits were thought to be more abundant in the past when, under low sulphate conditions, the formation of large amounts of iron oxides resulted in the deposition of banded iron formations^{4,5}. Under these circumstances, phosphorus removal by iron oxide adsorption could have been enhanced. Here we analyse the phosphorus and iron content of banded iron formations to show that ocean orthophosphate concentrations from

‡ Present address: Geological Institute, University of Copenhagen, Øster Voldgade 10, DK-1350 Copenhagen, Denmark.

3.2 to 1.9 billion years ago (during the Archaean and early Proterozoic eras) were probably only ~10–25% of present-day concentrations. We suggest therefore that low phosphorus availability should have significantly reduced rates of photosynthesis and carbon burial, thereby reducing the long-term oxygen production on the early Earth—as previously speculated⁴—and contributing to the low concentrations of atmospheric oxygen during the late Archaean and early Proterozoic.

The oxidation of dissolved ferrous iron produces insoluble iron oxyhydroxides, which strongly adsorb phosphates at pH values of less than 9 (ref. 6). The extent of phosphorus adsorption can be expressed as: $[P_{\text{ads}}] = K_{\text{ads}}[P_{\text{d}}][\text{Fe}^{3+}]$, where $[P_{\text{ads}}]$ is the concentration (μM) of phosphate adsorbed onto iron oxides, $[P_{\text{d}}]$ is the concentration (μM) of dissolved orthophosphate (PO_4^{3-}) in solution, $[\text{Fe}^{3+}]$ is the concentration (μM) of iron oxide particles, and K_{ads} (μM^{-1}) is the adsorption constant. Particles of iron oxide formed from Fe^{2+} oxidation in submarine hydrothermal systems adsorb phosphate with a K_{ads} value of $0.07 \pm 0.01 \mu\text{M}^{-1}$ (ref. 7). As phosphorus adsorbs predictably onto newly formed iron oxide surfaces, the phosphorus content of ancient sediments rich in iron oxide indicates, in principle, the phosphorus concentration of the water from which the oxides formed⁷.

Here we use the phosphorus and Fe^{3+} content of Archaean and early Proterozoic banded iron formations (BIFs), and the value for K_{ads} determined for modern hydrothermal systems⁷, to estimate the

$[P_{\text{d}}]$ of contemporaneous sea water:

$$[P_{\text{d}}] = (1/K_{\text{ads}})(P_{\text{ads}}/\text{Fe}^{3+}) \quad (1)$$

The ratio $P_{\text{ads}}/\text{Fe}^{3+}$ is directly available, in many instances, from BIF analyses (Fig. 1). An average Archaean and early Proterozoic phosphorus concentration of $0.15 \pm 0.15 \mu\text{M}$ is calculated (ranging from 0.03 to $0.29 \mu\text{M}$) for the waters where BIFs deposited (Fig. 1). Generally lower values of $[P_{\text{d}}]$ are indicated at 1.9–2.0 Gyr ago, compared to earlier BIFs.

We have avoided siderite-rich BIFs in the analysis as some contain unusual ^{13}C -depleted carbonates ($\delta^{13}\text{C} < -5\text{‰}$)^{8,9}, indicating the oxidation of sedimentary organic carbon, possibly by Fe reduction¹⁰, and the probable inclusion of original organic P into the total P pool. This would complicate our calculation. Furthermore, iron oxide (re)crystallization or iron reduction during early diagenesis could have induced a loss of P from the oxide-rich BIFs. In modern sediments preserving hydrothermally derived iron oxides, a loss of up to 50% of the originally adsorbed P is indicated¹¹. A 50% loss in P_{ads} would mean that our calculated $[P_{\text{d}}]$ could be too low by as much as a factor of two (ref. 11). However, this is probably a maximum correction, because hydrothermal sediments with higher iron content, approaching that of BIFs, have experienced ~25% P loss¹¹. Acknowledging that further tests of our approach are required, we conservatively estimate that BIFs deposited in sea water containing between, on average, 0.15 and $0.6 \mu\text{M}$ dissolved P.

We note that P retention in sediments under anoxic ocean conditions in the Archaean and early Proterozoic was probably different from today. This is indicated by the preservation of Fe oxides in BIFs, which deposited from anoxic ocean waters containing dissolved Fe^{2+} (ref. 4). Today, ocean anoxia results in a sulphidic water column which effectively reacts all of the available reactive Fe to form Fe sulphide minerals⁵. Thus, today, anoxia leads to some P loss from sediments^{12–15}, whereas anoxia during BIF deposition probably led to P retention on Fe oxides, and as ferrous phosphate minerals when there was Fe reduction in the sediment. The main difference between now and then was probably low sulphate concentrations in Archaean and early Proterozoic oceans, significantly reducing rates of sulphate reduction⁵. Even in the modern world, controlled sediment incubation experiments show that loss of adsorbed P is minimal under sediment diagenesis with low

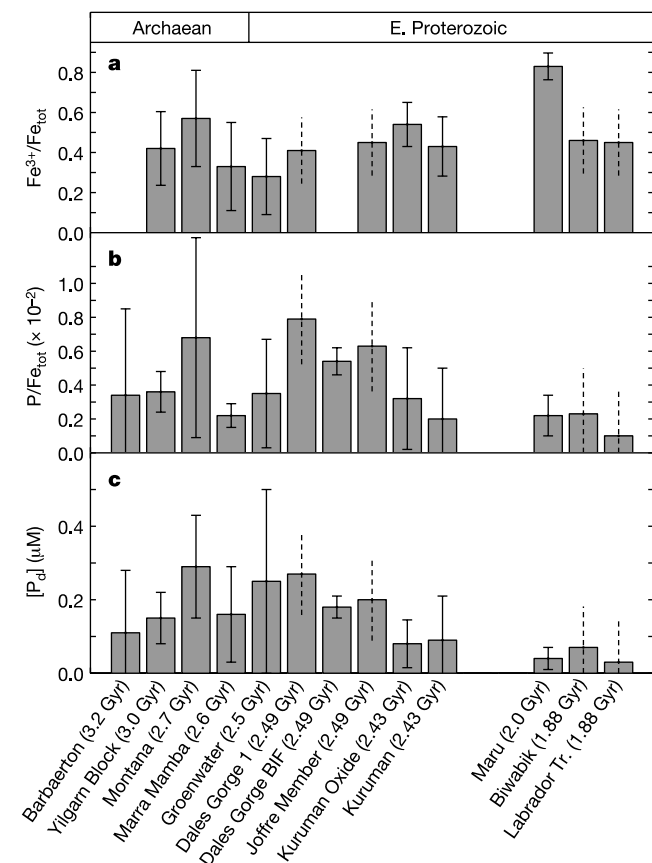


Figure 1 Element ratios in BIFs and calculated dissolved phosphate concentrations. Mole ratios of $\text{Fe}^{3+}/\text{Fe}_{\text{tot}}$ (**a**) and $\text{P}/\text{Fe}_{\text{tot}}$ (**b**) are from the literature (see Supplementary Information). **c**, Dissolved phosphate concentrations are calculated from equation (1). Phosphate loss during early diagenesis can result in up to a factor of two error in P_{d} (see text and ref. 11). Standard deviation for each BIF unit is shown as solid error bars. Dashed error bars indicate that no standard deviation could be calculated from the literature. In cases where Fe^{3+} content is not available, we use the mean $\text{Fe}^{3+}/\text{Fe}_{\text{tot}}$ ratio of 0.43.

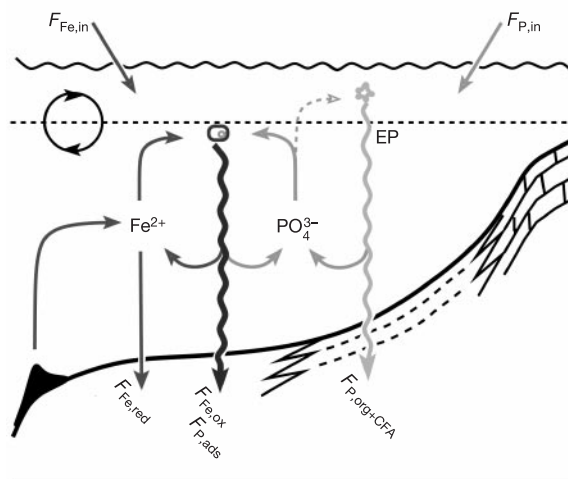


Figure 2 Simplified phosphate and iron cycle model of the Archaean and early Proterozoic. Dissolved ferrous iron (Fe^{2+}) is oxidized at the base of the oxic mixed layer, leading to iron oxide burial ($F_{\text{Fe,ox}} = \alpha F_{\text{Fe,in}}$) and BIFs with adsorbed P ($F_{\text{P,ads}}$). Organic matter export production (EP) is limited by upwelled P and leads to the burial of organically derived phosphate ($F_{\text{P,org+CFA}}$). For simplicity, ferrous iron is assumed to be buried ($F_{\text{Fe,red}}$) without associated phosphate. The total reactive iron input, $F_{\text{Fe,in}}$, is the sum of iron from land (dissolved+particulate) and hydrothermal iron.

sulphate concentrations¹⁶. The fact that BIFs are preserved, and contain phosphate, shows that the usual view of anoxia leading to P release is probably not valid during the periods in Earth's history when BIFs formed.

BIFs deposited in environments ranging from shelf and upper slope to the abyssal plain¹⁷. The Barberton BIF, associated with, and probably related to, local volcanism, was deposited at a water depth of approximately 900 m, the deepest water yet recognized for BIFs¹⁸. The other generally larger BIFs analysed here, lacking an obvious local volcanic source, deposited on the outer shelf or slope well below storm wave influence¹⁹. The Fe-oxide-rich facies of the Kuruman Iron Formation, a reasonable representative of the large 2.6-Gyr and younger BIFs in the data set, was deposited in deeper

water than the contemporaneous siderite-enriched BIF and shales¹⁹. Furthermore, the siderite-enriched BIF formed in water with dissolved inorganic carbon highly ¹³C-depleted compared to surface ocean water^{8,20}. Thus, the siderite-enriched BIF, and by extension the Fe-oxide-rich facies, deposited well below a pronounced chemocline where the accumulation of phosphate should have accompanied the steep gradients in isotopic composition of dissolved inorganic carbon. Therefore, our calculated phosphate concentrations characterize a region well below the chemocline and may represent the deep ocean. Phosphate concentrations of 0.15 to 0.6 μM are considerably less than the average modern ocean value of 2.3 μM .

Low phosphate concentrations in the Archaean and early Proterozoic oceans could have arisen from changes in the weathering flux of dissolved and reactive phosphate ($F_{\text{P,in}}$) to the ocean and/or significant changes in the ocean phosphate sinks (Fig. 2). We have no evidence supporting significantly reduced weathering fluxes of reactive phosphate. Continents may have reached their present size early in Earth's history and the intensity of chemical weathering, controlled by temperature and soil pH, might have been high in the Archaean and early Proterozoic. Alternatively, and more probably, low P concentrations originated from a strong P sink owing to significant adsorption onto, and removal by, iron oxide particles.

We use a simple ocean model to explore the possible influence of BIF deposition on ocean P concentrations (Fig. 2). P is delivered to the oceans from rivers ($F_{\text{P,in}}$) and is removed by adsorption onto iron oxides (P_{ads}) and as organically derived reactive P ($P_{\text{org+CFA}}$), which includes organic-bound P and P liberated from organic matter and mainly reprecipitated into carbonate fluorapatite (CFA) (refs 12, 13). Of the total reactive iron flux ($F_{\text{Fe,in}}$) to the ocean⁵, only a fraction (α) is buried as Fe oxides with adsorbed P; the rest is assumed to be buried as ferrous iron phases to which P does not adsorb (that is, sulphides and siderite). The burial flux of adsorbed P is then:

$$F_{\text{P,ads}} = (K_{\text{ads}}[P_{\text{d}}])(\alpha F_{\text{Fe,in}}) \quad (2)$$

The burial flux of $P_{\text{org+CFA}}$ (designated as $F_{\text{P,org+CFA}}$) depends on $[P_{\text{d}}]$, the upwelling velocity, u , of deep water supplying P_{d} to the euphotic zone, γ , and the relationship, η , between the export of organic P from the surface ocean and its burial^{4,21}:

$$F_{\text{P,org+CFA}} = \gamma(u[P_{\text{d}}]) \quad (3)$$

At steady state $F_{\text{P,in}} - (F_{\text{P,org+CFA}} + F_{\text{P,ads}}) = 0$, allowing a solution for $[P_{\text{d}}]$:

$$[P_{\text{d}}] = \frac{F_{\text{P,in}}}{\gamma u + \alpha K_{\text{ads}} F_{\text{Fe,in}}} \quad (4)$$

Figure 3 shows $[P_{\text{d}}]$ contoured as a function of the fraction, α , of total reactive iron buried with adsorbed P, versus reactive phosphate input ($F_{\text{P,in}}$). $[P_{\text{d}}]$ increases as $F_{\text{P,in}}$ increases, and decreases with increasing α . With today's range in $F_{\text{P,in}}$ (ref. 12) and $F_{\text{Fe,in}}$ (ref. 5) we find a $[P_{\text{d}}]$ similar to that obtained in the above BIF analysis if approximately half of the reactive iron flux is buried with adsorbed P ($0.4 < \alpha < 0.6$). Thus, low values of ocean P_{d} can be maintained if about half of $F_{\text{Fe,in}}$ is removed as freshly precipitated iron oxides. Such a high $\text{Fe}^{3+}/\text{Fe}_{\text{tot}}$ ratio is indicated for BIFs^{8,19} (Fig. 1a), and if this ratio represents the removal of $F_{\text{Fe,in}}$ in general, then the initial adsorption of P into BIF explains the low $[P_{\text{d}}]$ inferred for Archaean and early Proterozoic sea water.

Low values of ocean $[P_{\text{d}}]$ would have probably limited rates of primary production, and by extension, organic carbon burial, and the input of oxygen to the atmosphere^{4,20,22}. The relationship between $[P_{\text{d}}]$ and the burial rate of organic carbon ($F_{\text{C,org}}$) is given as the product of equation (4) and the ratio of organic carbon to organically derived P ($\eta_{\text{cp}} = C_{\text{org}}/P_{\text{org+CFA}}$):

$$F_{\text{C,org}} = \gamma(u[P_{\text{d}}])\eta_{\text{cp}} \quad (5)$$

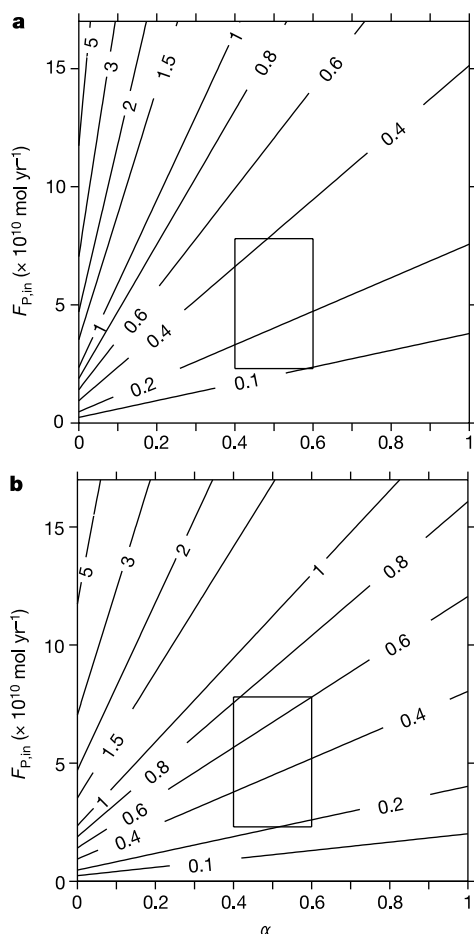


Figure 3 Modelled mean ocean phosphate concentration $[P_{\text{d}}]$ from equation (4). **a**, The standard case with $K_{\text{ads}} = 0.07 \mu\text{M}^{-1}$; **b**, The possible case of a 50% P_{ads} loss during early diagenesis with $K_{\text{ads}} \approx 0.035 \mu\text{M}^{-1}$. The $[P_{\text{d}}]$ values (in μM) are shown on the contours as a function of the fraction (α) of total iron buried with adsorbed P, and reactive phosphate input ($F_{\text{P,in}}$). Very low ocean phosphate concentrations occur if a large fraction of reactive iron input is buried with adsorbed P. Boxed regions correspond to the diagnosed Archaean conditions with the pre-industrial range of $F_{\text{P,in}}$ (refs 12, 30) and present $F_{\text{Fe,in}}$ (ref. 5). At present P, burial fluxes after early diagenesis are partitioned into: $F_{\text{P,org}} = 3.6 \times 10^{10} \text{ mol P yr}^{-1}$, $F_{\text{P,CFA}} = 3.6 \times 10^{10} \text{ mol P yr}^{-1}$ and $F_{\text{P,ads}} = 0.7 \times 10^{10} \text{ mol P yr}^{-1}$ (refs 12, 30). Part of the $F_{\text{P,CFA}}$ is originally reprecipitated from P liberated from decaying organic matter as well as P from reduction of iron oxides. Assuming, conservatively, that ~50% of the CFA burial flux originates from decay of organic matter, then $F_{\text{P,org+CFA}} = 5.4 \times 10^{10} \text{ mol P yr}^{-1}$. With a burial efficiency $\gamma = 1.9 \times 10^{-2}$, the above $F_{\text{P,org+CFA}}$ is reproduced for the present mean ocean P concentration (2.3 μM) and mixing $u = 1.26 \times 10^{18} \text{ l yr}^{-1}$. At present $F_{\text{P,org+CFA}}$ is associated with an organic carbon burial of $\sim 1 \times 10^{13} \text{ mol P yr}^{-1}$ (ref. 23), which gives a $C_{\text{org}}/P_{\text{org+CFA}}$ ratio of 185.

With values of γ , u and η_{CIP} comparable to today, a reduction in $[P_d]$ to 10–25% of present-day concentration (as we infer for the late Archaean and early Proterozoic) implies a 75–90% reduction in the rate of organic carbon burial. By contrast, if much less $P_{\text{org+CFA}}$ was buried with organic carbon, as has been argued for modern euxinic sediments¹⁴ (see also in ref. 15), then low $[P_d]$ would support substantial organic carbon burial, and oxygen production⁴. Thus, with $\eta_{\text{CIP}} \approx 1,000$, carbon burial would approach modern values²³. However, we believe that such high η_{CIP} ratios are probably inappropriate for an anoxic Fe-containing Precambrian ocean for two reasons. First, the high η_{CIP} ratios in modern euxinic sediments are partly a result of P loss^{13–15} under high sulphate conditions that were not present during BIF deposition²⁴. Second, even in modern settings, η_{CIP} may be much lower than previously thought¹⁵, with ratios of 150–200 recently reported from the euxinic, and rapidly depositing, Saanich Inlet sediments¹³. Thus, low $[P_d]$ in late Archaean and early Proterozoic oceans should have resulted in reduced rates of carbon burial. Higher heat flow and tectonic activity could have accelerated rates of geochemical cycling including rates of all of the processes discussed here, although the quantitative influence of heat flow on geochemical cycling rates is poorly known^{25,26}.

Previous analysis of the marine isotope record of organic and inorganic carbon suggests that for the Archaean (>2.5 Gyr ago) around 11% of the total carbon buried in ocean sediments was removed as organic carbon²⁷. By comparison, today, organic carbon represents 20% of the total carbon removal²⁷, or a burial percentage about twice that in the Archaean. Our results are generally consistent with this, and suggest reduced rates of organic carbon burial in the late Archaean and early Proterozoic. Further modelling studies, and a better resolution of the carbon isotope record, would better help to establish the concordance between our reconstruction of late Archaean and early Proterozoic nutrient chemistry and the carbon burial history as revealed from carbon isotope studies.

In addition to indicating relatively reduced rates of organic carbon burial in the Archaean, the isotope record of marine carbonates demonstrates at least one, and perhaps several, very large positive isotope excursions between 2.4 and 2.0 Gyr ago (refs 20, 27). These excursions indicate large increases in the relative proportions of organic carbon burial in ocean sediments, and possibly also indicate increases in the rate of organic carbon deposition²⁷. Recent compilations of the timing of BIF deposition indicate a curious lack of BIFs during this same time from 2.4 to around 2.0 Gyr ago (ref. 28). Furthermore, the isotope record of sedimentary sulphides²⁴ shows the first occurrence of highly ³⁴S-depleted marine sulphides around 2.4 Gyr ago. The sulphur isotope record indicates an increase in seawater sulphate concentrations to ~1 mM at around 2.4 Gyr ago, and a consequent increase in rates of sulphate reduction due to more sulphate availability²⁴. We propose that the burial pulse of organic carbon between 2.4 and 2.0 Gyr ago could have been driven, at least in part, by the P made available by increasing rates of sulphate reduction. Therefore, the rise in atmospheric oxygen concentrations^{4,20} that apparently accompanied the burial pulse of organic carbon²⁷ could also, in part, have resulted from a reduction in the Fe oxide sink and a consequently larger availability of P.

The decline of BIFs between 2.4 and 2.0 Gyr ago (ref. 28) could have been a logical consequence of the titration of deep-ocean Fe with sulphide owing to increasing rates of sulphate reduction⁵. For some reason not yet clear, BIF deposition, decreased organic carbon burial, and low P concentrations were reinitiated around 2.0 Gyr ago, persisting until around 1.8 Gyr ago (Fig. 1). Thereafter, consistent with an earlier proposal⁵, sulphidic conditions reoccurred. There is no indication of high burial fluxes of organic carbon after 1.8 Gyr ago, even though the iron oxide sink for phosphorus would have been substantially reduced under sulphidic ocean conditions.

Another nutrient may then have limited ocean primary productivity²⁹. □

Received 23 August 2001; accepted 26 March 2002.

- Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. E. Archaean molecular fossils and the early rise of eukaryotes. *Science* **285**, 1033–1036 (1999).
- Berner, R. A. Phosphate removal from sea water by adsorption on volcanogenic ferric oxides. *Earth Planet. Sci. Lett.* **18**, 77–86 (1973).
- Wheat, C. G., Feely, R. A. & Mottl, M. J. Phosphate removal by oceanic hydrothermal processes: An update of the phosphorus budget in the oceans. *Geochim. Cosmochim. Acta* **19**, 3593–3608 (1996).
- Holland, H. D. *The Chemical Evolution of the Atmosphere and Oceans* (Princeton Univ. Press, New York, 1984).
- Canfield, D. E. A new model for Proterozoic ocean chemistry. *Nature* **396**, 450–453 (1998).
- Stumm, W. & Morgan, J. J. *Aquatic Chemistry: An Introduction Emphasizing Chemical Equilibria in Natural Waters* (Wiley & Sons, New York, 1981).
- Feely, R. A., Trefry, J. H., Lebon, G. T. & German, C. R. The relationship between P/Fe and V/Fe ratios in hydrothermal precipitates and dissolved phosphate in seawater. *Geophys. Res. Lett.* **25**, 2253–2256 (1998).
- Beukes, N. J., Klein, C., Kaufman, A. J. & Hayes, J. M. Carbonate petrography, kerogen distribution, and carbon and oxygen isotope variations in an early Proterozoic transition from limestone to iron-formation deposition, Transvaal Supergroup, South Africa. *Econ. Geol.* **85**, 663–690 (1990).
- Kaufman, A. J., Hayes, J. M. & Klein, C. Primary and diagenetic controls of isotopic compositions of iron-formation carbonates. *Geochim. Cosmochim. Acta* **54**, 3461–3473 (1990).
- Walker, J. C. G. Suboxic diagenesis in banded iron formations. *Nature* **309**, 340–342 (1984).
- Schaller, T., Morford, J., Emerson, S. R. & Feely, R. A. Oxyanions in metalliferous sediments: Tracers for paleoseawater metal concentrations. *Geochim. Cosmochim. Acta* **63**, 2243–2254 (2000).
- Compton, J. et al. in *Marine Authigenesis: From Global to Microbial* (eds Glenn, C. R., Prevot, L. L. & Lucas, J.) 21–33 (Society for Sedimentary Geology, Tulsa, 2000).
- Filippelli, G. M. Carbon and phosphorus cycling in anoxic sediments of the Saanich Inlet, British Columbia. *Mar. Geol.* **174**, 307–321 (2001).
- Van Cappellen, P. & Ingall, E. D. Redox stabilization of the atmosphere and oceans by phosphorus-limited marine productivity. *Science* **271**, 493–496 (1996).
- Anderson, L. D., Delaney, M. L. & Faul, K. L. Carbon to phosphorus ratios in sediments: Implications for nutrient cycling. *Glob. Biogeochem. Cycles* **15**, 65–79 (2001).
- Roden, E. E. & Edmonds, J. W. Phosphate mobilization in iron-rich anaerobic sediments: Microbial Fe(III) oxide reduction versus iron-sulfide formation. *Archiv. Hydrobiol.* **139**, 347–378 (1997).
- Fralick, P. & Barrett, T. J. in *Sedimentary Facies Analysis: A Tribute to the Research and Teaching of Harold G. Reading* (ed. Plint, A. G.) 137–156 (Special Publication of the International Association of Sedimentologists, Oxford, 1995).
- de Ronde, C. E. J., Channer, D. M., De, R., Faure, K., Bray, C. J. & Spooner, E. T. C. Fluid chemistry of Archaean seafloor hydrothermal vents: Implication for the composition of circa 3.2 Ga seawater. *Geochim. Cosmochim. Acta* **61**, 4025–4042 (1997).
- Klein, C. & Beukes, N. J. Geochemistry and sedimentology of a facies transition from limestone to iron-formation deposition in the early Proterozoic Transvaal Supergroup, South Africa. *Econ. Geol.* **84**, 1733–1774 (1989).
- Karhu, J. A. & Holland, H. D. Carbon isotopes and the rise of atmospheric oxygen. *Geology* **24**, 867–870 (1996).
- Shaffer, G. A model of biogeochemical cycling of phosphorus, nitrogen, oxygen, and sulphur in the ocean: One step toward a global climate model. *J. Geophys. Res.* **94**, 1979–2004 (1989).
- Lenton, T. M. & Watson, A. J. Redfield revisited 2. What regulates the oxygen content of the atmosphere? *Glob. Biogeochem. Cycles* **14**, 249–268 (2000).
- Holland, H. D. *The Chemistry of the Atmosphere and Oceans* (Wiley, New York, 1978).
- Canfield, D. E., Habicht, K. S. & Thamdrup, B. The archaean sulfur cycle and the early history of atmospheric oxygen. *Science* **288**, 658–661 (2000).
- Des Marais, D. J. Isotopic evolution of the biogeochemical carbon cycle during the Proterozoic eon. *Org. Geochem.* **27**, 185–193 (1997).
- Sleep, N. H. & Zahnle, K. Carbon dioxide cycling and implications for climate on ancient Earth. *J. Geophys. Res.* **106**, 1373–1399 (2001).
- Des Marais, D. J., Strauss, H., Summons, R. E. & Hayes, J. M. Carbon isotope evidence from the stepwise oxidation of the Proterozoic environment. *Nature* **359**, 605–609 (1992).
- Isley, A. E. & Abbott, D. H. Plume-related mafic volcanism and the deposition of banded iron formation. *J. Geophys. Res.* **104**, 15461–15477 (1999).
- Anbar, A. D. & Knoll, A. H. Trace metal limitation of primary production 1.85–1.25 Ga. (American Geophysical Union Fall Meeting, 1999); available at <http://www.agu.org/meetings/fm99top.html>.
- Berner, E. K. & Berner, R. A. *The Global Water Cycle, Geochemistry and Environment* (Prentice Hall, New Jersey, 1994).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank J. Hayes and T. Lenton for comments and suggestions. This work was funded by the Danish National Research Foundation (Danmarks Grundforskningsfond) and the Danish National Science Foundation (SNF).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.J.B. (e-mail: cjb@geo.geol.ku.dk).

Vestibular evidence for the evolution of aquatic behaviour in early cetaceans

F. Spoor*, S. Bajpai†, S. T. Hussain‡, K. Kumar§ & J. G. M. Thewissen||

* Department of Anatomy & Developmental Biology, University College London, Rockefeller Building, University Street, London WC1E 6JJ, UK

† Department of Earth Sciences, Indian Institute of Technology, Roorkee 247 667, India

‡ Department of Anatomy, College of Medicine, Howard University, Washington DC 20059, USA

§ Wadia Institute of Himalayan Geology, Dehradun 248 001, India

|| Department of Anatomy, Northeastern Ohio Universities College of Medicine, Rootstown, Ohio 44272, USA

Early cetaceans evolved from terrestrial quadrupeds to obligate swimmers, a change that is traditionally studied by functional analysis of the postcranial skeleton¹. Here we assess the evolution of cetacean locomotor behaviour from an independent perspective by looking at the semicircular canal system, one of the main sense organs involved in neural control of locomotion². Extant cetaceans are found to be unique in that their canal arc size, corrected for body mass, is approximately three times smaller than in other mammals. This reduces the sensitivity of the canal system, most plausibly to match the fast body rotations that characterize cetacean behaviour. Eocene fossils show that the new sensory regime, incompatible with terrestrial competence, developed quickly and early in cetacean evolution, as soon as the taxa are associated with marine environments. Dedicated agile swimming of cetaceans thus appeared to have originated as a rapid and fundamental shift in locomotion rather than as the gradual transition suggested by postcranial evidence. We hypothesize that the unparalleled modification of the semicircular canal system represented a key 'point of no return' event in early

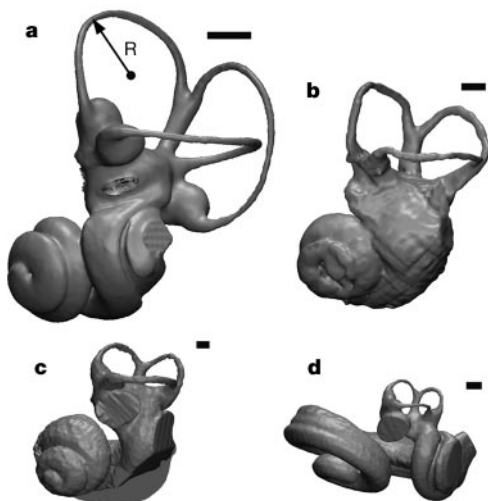


Figure 1 Lateral view of left bony labyrinths. **a**, *Galago moholi*, a particularly agile non-cetacean mammal; **b**, *Ichthyolestes pinfoldi*; **c**, *Indocetus ramani* (inferior part of basal turn partially reconstructed); **d**, extant *Tursiops truncatus*. The size of the labyrinths is corrected for body mass according to the mammalian regression given in Fig. 2a. Scale bars are 1 mm. The radius of curvature (*R*) of the anterior semicircular canal, expressing its arc size, is indicated for *Galago*. The apex of the cochlea faces inferiorly rather than laterally in *Tursiops*.

cetacean evolution, leading to full independence from life on land.

Early cetacean evolution, marked by the emergence of obligate aquatic behaviour, represents one of the major morphological shifts in the radiation of mammals. Modifications to the postcranial skeleton during this process are increasingly well-documented^{3–9}. Pakicetids, early Eocene basal cetaceans, were terrestrial quadrupeds with a long neck and cursorial limb morphology⁹. By the late middle Eocene, obligate aquatic dorudontids approached modern cetaceans in body form, having a tail fluke, a strongly shortened neck, and near-absent hindlimbs¹⁰. Taxa which represent bridging nodes on the cladogram show intermediate morphologies, which have been inferred to correspond with otter-like swimming combined with varying degrees of terrestrial capability^{4–8,11}. Our knowledge of the behavioural changes that crucially must have driven the postcranial adaptations is based on functional analysis of the affected morphology itself. This approach is marred by the difficulty of recognizing whether plesiomorphic morphologies were either simply retained from the ancestor, or actively selected for in their original or a newly acquired functional role. We therefore sought to assess early cetacean locomotor behaviour using an alternative and independent source, by examining the semicircular canal system, part of the vestibular apparatus (organ of balance) housed in the

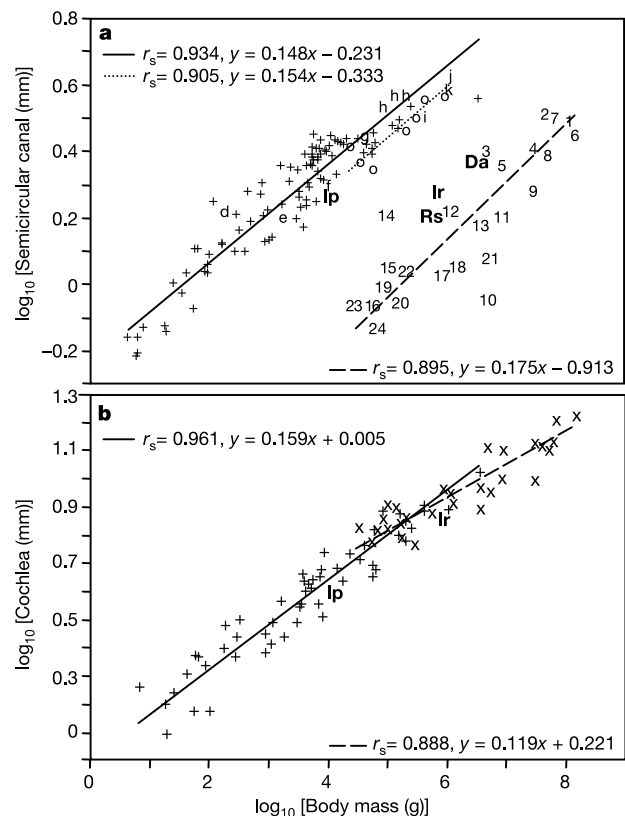


Figure 2 Relationship of semicircular canal and cochlea sizes to body mass. **a**, **b**, Bivariate double logarithmic plots between body mass and the mean radius of curvature of the three semicircular canals (**a**), and the size of the cochlea (**b**). Codes for cetacean species are given in Table 1. Data points are labelled as follows: x, extant cetaceans; o, artiodactyls; crosses, other mammal species; d, *Galago moholi* (shown in Fig. 1a); e, *Ornithorhynchus anatinus*; f, *Lutra lutra*; g, *Hydrochaeris hydrochaeris*; h, Phocidae; i, *Dugong dugon*; j, *Hippopotamus amphibius*; k, *Odobenus rosmarus*; e–k, amphibious or aquatic non-cetaceans. The Spearman rank correlation coefficients (r_s) and reduced major axis regressions are given for the extant cetaceans (dashed line), the non-cetacean mammals (solid line) and the artiodactyls (dotted line); **a**,

inner ear. It records angular motion of the head and contributes, through reflex mechanisms, to the sensory control of locomotion, including the stabilization of the head and eyes². The arc size of semicircular canals is a determinant of their sensitivity^{12,13} and has been found to correlate with the degree of locomotor agility^{14,15}.

Cetacean semicircular canals have been noted as small in size^{16–23}, but this phenomenon has been studied neither quantitatively nor systematically. We therefore compared the canal sizes of 24 extant cetaceans and 106 other mammalian species, representing all major

Table 1 The cetacean species and other mammalian orders analysed

	Semicircular canal	Cochlea	Code
Monotremata	1	–	
Didelphimorphia	3	3	
Notoryctemorphia	1	1	
Dasyuromorphia	3	3	
Peramelemorphia	1	1	
Insectivora	2	2	
Chiroptera	7	2	
Scandentia	1	–	
Primates	43	11	
Xenarthra	3	2	
Lagomorpha	2	2	
Rodentia	12	6	
Carnivora (Fissipedia)	12	10	
Pinnipedia	4	3	
Proboscidea	1	1	
Sirenia	1	1	
Perissodactyla	1	1	
Artiodactyla	8	5	
Balaenidae			
<i>Balaena mysticetus</i>	1	–	1
<i>Eubalaena glacialis</i>	2	2	2
Neobalaenidae			
<i>Caperea marginata</i>	1	1	3
Eschrichtiidae			
<i>Eschrichtius robustus</i>	1	1	4
Balaenopteridae			
<i>Balaenoptera acutorostrata</i>	1	1	5
<i>Balaenoptera borealis</i>	–	1	–
<i>Balaenoptera musculus</i>	1	1	6
<i>Balaenoptera physalis</i>	1	1	7
<i>Megaptera novaeangliae</i>	2	2	8
Physeteridae			
<i>Physeter catodon</i>	1	1	9
Kogiidae			
<i>Kogia sp.</i>	1	1	10
Ziphiidae			
<i>Berardius bairdi</i>	1	1	11
<i>Mesoplodon densirostris</i>	1	1	12
<i>Ziphius cavirostris</i>	1	1	13
Platanistidae			
<i>Platanista gangetica</i>	1	1	14
Iniidae			
<i>Inia geoffrensis</i>	1	1	15
Pontoporiidae			
<i>Pontoporia blainvillei</i>	1	1	16
Monodontidae			
<i>Delphinapterus leucas</i>	1	1	17
<i>Monodon monoceros</i>	1	1	18
Delphinidae			
<i>Delphinus sp.</i>	1	2	19
<i>Feresa attenuata</i>	–	1	–
<i>Globicephala sp.</i>	–	1	–
<i>Grampus griseus</i>	–	1	–
<i>Lagenorhynchus obliquidens</i>	1	1	20
<i>Orcinus orca</i>	1	1	21
<i>Stenalla sp.</i>	–	1	–
<i>Tursiops truncatus</i>	3	3	22
Phocoenidae			
<i>Neophocaena phocaenoides</i>	1	1	23
<i>Phocoena phocoena</i>	1	1	24
Pakicetidae			
<i>Ichthyolestes pinfoldi</i> HGSP96434	1	1	Ip
Remingtonocetidae			
<i>Remingtonocetus sp.</i> RUSB2529	1	–	Rs
Protocetidae			
<i>Indocetus ramani</i> LUVF11034	1	1	Ir
Dorudontidae			
<i>Dorudon atrox</i> BM(NH)-M9266	1	–	Da

The number of specimens are given for the cetaceans, and the number of species for the other mammalian orders, listed separately for the semicircular canal and cochlea measurements. The codes used in Fig. 1 are indicated for the cetaceans. A full list of the measurements for all species is available as Supplementary Information.

orders (Table 1; Fig. 1a, d; Supplementary Information). In cetaceans the canal arc size scales with body mass with similar, strong negative allometry, as it does in other mammals (Fig. 2a). However, the canal arcs are, on average, three times smaller than those of other mammals of the same body mass, without overlap between the respective ranges of interspecific variation. Hence, the semicircular canals of the blue whale (*Balaenoptera musculus*) are just smaller than the average human size, and those of the bottle-nosed dolphin (*Tursiops truncatus*) are considerably smaller than in the brown rat. We also measured the semicircular canal size of four Eocene cetaceans. They represent respective nodes on the cetacean phylogenetic tree, and are: *Ichthyolestes*, a pakicetid from Pakistan; *Remingtonocetus*, a remingtonocetid from India; *Indocetus*, a protocetid from India; and *Dorudon*, a dorudontid from Egypt (Figs 1b, c and 3). The oldest of these, *Ichthyolestes*, shows a canal size similar to that of the closest living relatives of cetaceans, the artiodactyls^{8,9} (Fig. 2a). In contrast, *Remingtonocetus*, *Indocetus* and *Dorudon* all have small canal sizes that fall at the upper margin of the range of extant cetaceans (Fig. 2a). The size of the cochlea in extant and fossil cetaceans is close to that in other mammals (Fig. 2b), which confirms that the unprecedented size reduction specifically concerns the semicircular canals rather than the entire cetacean inner ear.

Functionally, the small semicircular canal size of cetaceans has been interpreted as a vestigial condition, in which receptors are lost because of limited neck and eye motility^{14,20}. However, the observed pattern of allometric scaling in parallel with terrestrial mammals points at structured functional adjustment of the system rather than degeneration and redundancy. Central to understanding the reduced canal size is that extant cetaceans, freed from the restrictions of gravitational pull and the need for substrate contact, are particularly acrobatic when compared with terrestrial animals of similar body size. Paradoxically, among non-cetacean mammals acrobatic species show canals with a relatively large arc^{14,15} (Fig. 1a). This increases the canals' sensitivity^{12,13}, and thus the ability to resolve small changes in angular head motion¹⁵. For animals showing fast and highly manoeuvrable locomotion such accurate sensory information is vital to maintain body coordination. Importantly, the canals of such species can be sensitive, without the risk of constant overstimulation, because they are part of a feedback system. The canals supply the vestibulo-collic reflex, which stabilizes the head by compensatory neck movements, thus keeping the input signal of the canals within limits. In extant cetaceans, however, reflex stabilization of the head is ineffective, because their streamlined bodies, marked by shortened and frequently fused cervical vertebrae, allow very little neck motility. The most plausible reason

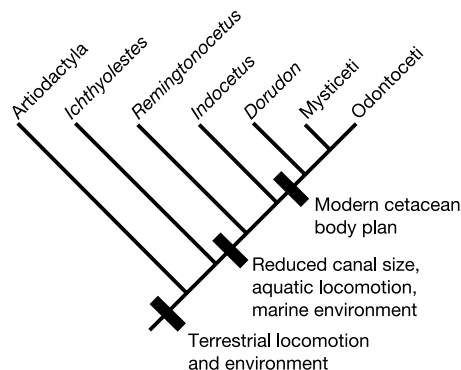


Figure 3 Cladogram showing the phylogenetic relationship of Eocene cetacean taxa examined here, their sister group the artiodactyls and the modern mysticetes and odontocetes. Key states of morphology, locomotion and environment are indicated. Modern cetacean bodyplan refers to, among other things, the total loss of hindlimb locomotor function and the presence of a tail fluke.

for the small arc size of their semicircular canals is thus to reduce sensitivity to match the high levels of uncompensated angular motion. In an aquatic environment this reduction is arguably less critical than in, for example, an arboreal setting where less accurate sensory clues easily impair locomotor control. The canal sizes of amphibious and aquatic, non-cetacean mammals in our sample are within the range for terrestrial species (Fig. 2a). However, none of these combine great aquatic agility with lack of neck motility.

The first appearance of small semicircular canals in the middle Eocene genera *Remingtonocetus* and *Indocetus* shows that the reduction in canal size took place rapidly and early in cetacean evolution, within five million years of the origin of the order. The palaeoenvironmental context of the fossils²⁴ demonstrates that this event accompanied the invasion of marine environments (Fig. 3). The newly evolved and highly derived vestibular sensory regime was almost certainly incompatible with any terrestrial locomotion beyond cautious beach-bound crawling, which indicates that dedicated agile swimming of cetaceans originated as a rapid and fundamental shift in behaviour. As such, the modification of the semicircular canal system represented a crucial 'point of no return' event in early cetacean evolution, which excluded a prolonged semi-aquatic phase. These findings strongly contrast with the pattern of gradual change shown by the postcranial skeleton, with the near-modern body plan and fluke-based propulsion of dorudontids emerging eight million years after the canal system modified (Fig. 3). In that period the well-developed hindlimbs of remingtonocetids and protocetids^{5,7,8} had a newly acquired function in powerful swimming, with some evidence pointing at a persisting degree of weight-bearing²⁵.

There remains the question of why *Remingtonocetus* and *Indocetus* would have shown increased levels of angular head motion, as is suggested by their reduced semicircular canal size. Although their neck length was shortened compared with pakicetids and other cursorial species^{9,26,27}, this was not to the extent seen in extant cetaceans. We hypothesize that the key underlying factor is the large, remarkably long-snouted head of Eocene cetaceans^{5,7,8}, including the oldest, terrestrial pakicetids⁹. When adapting to agile aquatic locomotion this particular morphology favours hydrodynamic integration of the head and trunk to reduce drag, minimizing neck movements at the cost of effective reflex stabilization of the head. The semicircular canals, as a functionally constrained sensory system, instantaneously adjusted to such a kinematic change, whereas skeletal modification leading to extreme shortening of the neck was a more gradual process.

Cetaceans are unusual in many ways, but well-known characters such as a relatively large brain, a tail fluke, echolocation, filterfeeding and mandibular hearing are all found in other mammals as well. The highly derived vestibular apparatus of cetaceans, on the other hand, is unique, and it is the first cetacean organ known to reach modern morphology. That early aquatic cetaceans had semicircular canals that do not resemble those of extant amphibious or aquatic non-cetacean mammals confirms that they behaved in ways that have no direct parallels among extant species. □

Methods

Size measurements of the semicircular canals and the cochlea were taken from computed tomography (CT) scans, plexiglas casts of the bony labyrinth, and the literature (Supplementary Information). CT scans, made with Siemens Somatom scanners at the National Museum of Natural History in Washington DC and the Hammersmith Hospital in London (*Dorudon atrox* only), have a pixel size of 0.1 mm, a slice thickness of 1.0 mm and a slice increment of 0.5 mm. CT scans of the other fossils and *Tursiops* were made with a Norland XCT Research M scanner, and these values are 0.07 mm, 0.50 mm and 0.25 mm, respectively. Scans were made coronally, and measurements were taken from planar reformatted images showing the full extent of each canal, using Voxblast 3.1 (Vaytek). Measurements of the casts were made using a binocular microscope. The literature sources used either give the required dimensions, list measurements from which they can be calculated, or include well-scaled images from which measurements can be taken.

The arc size of the semicircular canals was quantified by the radius of curvature, defined as half the average of the arc height and width (Fig. 1a; ref. 28). The results are

similar for the anterior, posterior and lateral canals so the mean of their three radii of curvature was used. The overall size of the cochlea was quantified by the mean of the 'slant height' and the diameters of the first and second turn. The slant height is the distance from the apex to the posteriormost edge of the round window¹⁷, and the diameters were measured inferosuperiorly or mediolaterally, depending on the orientation of the cochlea. Owing to the preservation of the fossils the cochlear measurements could only be taken for *Ichthyolestes* and *Indocetus*.

Body mass data, including the estimates for the fossil taxa, are listed in the Supplementary Information. The estimate for *Remingtonocetus* is based on the mean value reported for *Dalanistes*⁵, which is closely related, but is 20% larger, and the estimate for *Indocetus* is the value reported for closely related and similar-sized *Rodhocetus*. The values are based on vertebral centrum size²⁷. The body mass of *Dorudon* is estimated from the full skeletal length of two adult specimens²⁹. The estimate for *Ichthyolestes* is based on bones of several individuals⁹. A braincase (H-GSP 98134) is identical in bizygomatic width (98 mm) to an average individual of *Canis latrans*. Long bones of *Ichthyolestes* (humerus, H-GSP 30128; femur, H-GSP 30420; tibia, H-GSP 98189) are similar in length to those of *Vulpes vulpes*, but are more robust. The body mass value of *Ichthyolestes* is taken as the mean of these two canids. Even considerable changes to these estimates do not alter the conclusions of the analyses.

Grade shifts in comparative studies can be demonstrated by inspecting the phylogenetically independent contrast of the basal node. Here, this is unnecessary because the grade shift in arc size shown by the non-pakicetid cetacean semicircular canals is unambiguous, marked by complete separation from the artiodactyls and other mammals (Fig. 2a). Three-dimensional reconstructions of the labyrinths in Fig. 1 were made with Voxel-man (Univ. of Hamburg)³⁰.

Received 25 July 2001; accepted 30 January 2002.

1. Thewissen, J. G. M. in *The Emergence of Whales* (ed. Thewissen, J. G. M.) 451–464 (Plenum, New York, 1998).
2. Schwartz, D. W. F. & Tomlinson, R. D. in *Neurotology* (eds Jackler, R. K. & Brackmann, D. E.) 59–98 (Mosby, St Louis, 1994).
3. Gingerich, P. D., Smith, B. H. & Simons, E. L. Hind limbs of Eocene *Basilosaurus*: evidence of feet in whales. *Science* **249**, 154–157 (1990).
4. Thewissen, J. G. M., Hussain, S. T. & Arif, M. Fossil evidence for the origin of aquatic locomotion in archaeocetes. *Science* **263**, 210–212 (1994).
5. Gingerich, P. D., Arif, M., Clyde, W. C. New archaeocetes (Mammalia, Cetacea) from the middle Eocene Domanda Formation of the Sulaiman Range, Punjab (Pakistan). *Contrib. Mus. Paleontol. Univ. Michigan* **29**, 291–330 (1995).
6. Hulbert, R. C. in *The Emergence of Whales* (ed. Thewissen, J. G. M.) 235–267 (Plenum, New York, 1998).
7. Bajpai, S. & Thewissen, J. G. M. A new, diminutive Eocene whale from Kachchh (Gujarat, India) and its implications for locomotor evolution of cetaceans. *Curr. Sci.* **79**, 1478–1482 (2000).
8. Gingerich, P. D., Ul Haq, M., Zalmout, L. S., Khan, I. H. & Malkani, M. S. Origin of whales for early artiodactyls: hand and feet of Eocene Protocetidae from Pakistan. *Science* **293**, 2239–2242 (2001).
9. Thewissen, J. G. M., Williams, E. M. & Hussain, S. T. Skeletons of terrestrial cetaceans and the relationship of whales to artiodactyls. *Nature* **413**, 277–281 (2001).
10. Uhen, M. in *The Emergence of Whales* (ed. Thewissen, J. G. M.) 29–61 (Plenum, New York, 1998).
11. Thewissen, J. G. M. & Fish, F. E. Locomotor evolution in the earliest cetaceans: functional model, modern analogues, and paleontological evidence. *Paleobiology* **23**, 482–490 (1997).
12. Oman, C. M., Marcus, E. N. & Curthoys, I. S. The influence of the semicircular canal morphology on endolymph flow dynamics. *Acta Otolaryngol.* **103**, 1–13 (1987).
13. Muller, M. Semicircular duct dimensions and sensitivity of the vertebrate vestibular system. *J. Theor. Biol.* **167**, 239–256 (1994).
14. Spoor, F. & Zonneveld, F. Comparative review of the human bony labyrinth. *Yb. Phys. Anthropol.* **41**, 211–251 (1998).
15. Spoor, F. The semicircular canal system and locomotor behaviour, with special reference to hominin evolution. *Courier Forschungsinstitut Senckenberg* (in the press).
16. Hyrtl, J. *Vergleichend-anatomische Untersuchungen Über das innere Gehörorgan des Menschen und der Säugethiere* (Ehrlich, Prague, 1845).
17. Gray, A. A. *The Labyrinth of Animals* Vol. 2 (Churchill, London, 1908).
18. Yamada, M. & Yoshizaki, F. Osseous labyrinth of Cetacea. *Sci. Rep. Whale Res. Inst. (Tokyo)* **14**, 291–304 (1959).
19. Fleischer, G. Hearing in extinct cetaceans as determined by cochlear structure. *J. Paleontol.* **50**, 133–152 (1976).
20. Ketten, D. R. in *Marine Mammals Sensory Systems* (ed. Thomas, J.) 53–57 (Plenum, New York, 1992).
21. Luo, Z. & Eastman, E. R. Petrosal and inner ear of a squalodontoid whale: implications for evolution of hearing in odontocetes. *J. Vert. Paleontol.* **15**, 431–442 (1995).
22. Luo, Z. & Marsh, K. Petrosal (periotic) and inner ear of a Pliocene kogiine whale (*Kogiinae*, Odontoceti): implications on the relationships and hearing evolution of toothed whales. *J. Vert. Paleontol.* **16**, 328–348 (1996).
23. Geisler, J. H. & Luo, Z. The petrosal and inner ear of *Herpetocetus* sp. (Mammalia: Cetacea) and their implications for the phylogeny and hearing of archaic mysticetes. *J. Paleontol.* **70**, 1045–1066 (1996).
24. Williams, E. M. in *The Emergence of Whales* (ed. Thewissen, J. G. M.) 1–28 (Plenum, New York, 1998).
25. Madar, S. I. in *The Emergence of Whales* (ed. Thewissen, J. G. M.) 353–377 (Plenum, New York, 1998).
26. Buchholz, E. A. in *The Emergence of Whales* (ed. Thewissen, J. G. M.) 325–351 (Plenum, New York, 1998).
27. Gingerich, P. D. in *The Emergence of Whales* (ed. Thewissen, J. G. M.) 423–449 (Plenum, New York, 1998).
28. Spoor, C. F. & Zonneveld, F. W. Morphometry of the primate bony labyrinth: a new method based on high-resolution computed tomography. *J. Anat.* **186**, 271–286 (1995).
29. Marino, L. et al. Endocranial volume of mid-late Eocene archaeocetes (order: Cetacea) revealed by computed tomography: implications for cetacean brain evolution. *J. Mamm. Evol.* **7**, 81–94 (2000).
30. Hühne, K. H. et al. A new representation of knowledge concerning human anatomy and function. *Nature Med.* **1**, 506–511 (1995).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We thank E. Allen, H. Chatterjee, J. Hooker, J. Mead, C. Potter and T. Yamada for access to specimens, and R. Barton, E. Blum, M. Colbert, C. Dean, J. DePonte, B. Frohlich, K. Grecco, K. H. Höhne, N. Jeffery, B. Jonsdottir, W. Jungers, R. Ketcham, K. Kupczik, D. Lieberman, Z. Luo, J. Moore, M. Muller, J. Neiger, C. Pellow, D. Plummer, A. Pommert, C. Ross, G. B. Schneider and A. Walker for their help. This research was supported by grants for the UCL Graduate School to F.S. and from the NSF to J.G.M.T. and S.T.H.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.S. (e-mail: f.spoor@ucl.ac.uk).

Allometric cascade as a unifying principle of body mass effects on metabolism

Charles-A. Darveau*, Raul K. Suarez†, Russel D. Andrews* & Peter W. Hochachka*

* Department of Zoology, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada

† Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, California 93106-9610, USA

The power function of basal metabolic rate scaling is expressed as aM^b , where a corresponds to a scaling constant (intercept), M is body mass, and b is the scaling exponent. The $3/4$ power law (the best-fit b value for mammals) was developed from Kleiber's original analysis¹ and, since then, most workers have searched for a single cause to explain the observed allometry. Here we present a multiple-causes model of allometry, where the exponent b is the sum of the influences of multiple contributors to metabolism and control. The relative strength of each contributor, with its own characteristic exponent value, is determined by the control contribution. To illustrate its use, we apply this model to maximum versus basal metabolic rates to explain the differing scaling behaviour of these two biological states in mammals. The main difference in scaling is that, for the basal metabolic rate, the O_2 delivery steps contribute almost nothing to the global b scaling exponent, whereas for the maximum metabolic rate, the O_2 delivery steps significantly increase the global b value.

Until now, the classical approach to the basal metabolic rate (BMR) allometry problem has been to search for the single driving force or single rate-limiting step enforcing its scaling behaviour on overall metabolism. However, this concept in metabolic regulation studies was abandoned during the 1960s and was replaced by the concept of multiple control sites in metabolic pathways. By the 1980s and 1990s, rigorous mathematical models allowed control analysis to quantify the control contributions of different steps in metabolic pathways^{2–5} or in whole-organism physiological processes⁶. A key point, that control is vested in both energy supply and energy demand pathways^{3–5}, is important because many biologists estimate metabolic rate from O_2 uptake rates per gram per minute or from heat output rates. These parameter estimates are valid because they in turn are stoichiometrically related to complex pathways of adenosine triphosphate (ATP) synthesis rates and ATP utilization rates (however, see the discussion below on the 'O₂

wasting effect' of proton leak). In organisms at steady state the fluxes through ATP synthesis and utilization pathways are equal and mass-specific metabolic rate is equivalent to μmol s ATP cycled through ATP supply and ATP demand pathways per gram per minute. That is why it should not be surprising that modern metabolic studies identify regulatory roles in overall metabolic fluxes in both energy-supply and energy-demand pathways^{2–6}. In the traditional power function

$$\text{BMR} = aM^b \quad (1)$$

(a , M and b defined above), examination of the effects of body mass on metabolism should ideally consider the value of exponent b in equation (1) and the control contribution (or control coefficient) for each and every major step in ATP supply and ATP demand pathways in the system under investigation.

To sort out the varied contributions to control of net metabolic flux, say net O_2 flux, experimenters determine the fractional change in organismal O_2 flux caused by a fractional change in flux capacity through any given step or process in the path of O_2 from lungs to mitochondria in working tissues. The fractional change in organismal flux divided by the fractional change in capacity represents the control coefficient at each step^{2–6}. To illustrate, if the O_2 flux capacity of the lung is increased by 50% but only a 25% change in overall O_2 flux is achieved, the control coefficient for lung diffusion is 0.5, equal to the fractional change in overall O_2 flux divided by the fractional change in lung flux capacity. All the control coefficients in the pathway by definition add up to 1. In a system with a classical single rate-limiting step, say O_2 delivery^{7,8}, the control coefficient for that step would be essentially 1; that is, all control is vested in this process. The latter situation is never found in metabolic systems of any complexity^{3–6}. Under resting conditions all steps in the mammalian O_2 delivery chain necessarily display huge excess capacities and thus could not possibly be 'rate limiting' in the classical sense, while as we shall discuss further below, under maximum O_2 flux conditions, control is shared rather evenly among several major links in the chain. Hence the usually explicit, sometimes implicit, assumption of traditional allometry studies^{7,8}, that there is a single rate-limiting step or process that accounts for the b value in equation (1), is flawed.

We present here a multiple-causes model of allometry, in which there are multiple contributors to control, each with their own characteristic b values, that with their control contributions determine the value of the b scaling coefficient for overall energy metabolism. We can express this relationship as:

$$\text{MR} = a \sum c_i M^{b_i} \quad (2)$$

where MR is the metabolic rate in any given state, M is body mass, a is the intercept, b_i the scaling exponent of the process i , and c_i the control coefficient of the process i . (In most cases, a will be taken simply as the intercept, which is traditional in the allometry field. As ATP-utilizing processes are linked in parallel, each may have a unique a_i value; in sum these equal the baseline ATP turnover. On the energy supply side, the processes are linked in series with only one intercept at steady state indicating the same rate of baseline ATP turnover. For some purposes it is useful to decompose a into $a_{i,\text{max}}$ the flux capacity through step i , and f_i the fraction of $a_{i,\text{max}}$ used in a given metabolic state (see discussion below)).

Depending on metabolic state, the overall MR corresponds to the sum of the various contributors to ATP turnover, and the b exponent is the sum of the scaling exponents of all contributors, modulated by their control coefficients. In other words, each step in the linked series (that makes up the process, say O_2 flux, under study) possesses its own scaling behaviour, and its control coefficient c_i determines the degree to which its unique b_i exponent influences the global b value. For energy metabolism in mammals operating at varying rates between basal and maximum levels,

equation (2) applies to the sums of processes (such as ventilation, cardiac work, circulation) occurring at different sites in the body. At maximal sustainable work rates (equivalent to maximum sustainable ATP turnover rates), major contributors to control include the lung⁶, the heart and circulation⁹, cellular-level energy-supply pathways^{3–5}, which themselves have steps with low and high control contributions, and finally the ATP-demand pathways^{3–5}.

The concept of an allometric cascade arises from the layering of function at various levels of organization. The numerous steps involved in intracellular pathways of ATP demand and supply, each with its own characteristic b_i and c_i values, are at the base of the cascade. Other key steps, such as transporters, exchangers and pumps, function at the membrane-based cell surface, interfacing and integrating intra- and extracellular functions. Not all cells are identical of course, so another layering of function is found at organ and tissue levels of organization, where organ physiologies are integrated into organism BMR. Thus the overall scaling of BMR is a consequence of the interaction of scaling of these various 'functional units' involved in specific metabolic states. In short, we view this multi-cause model of scaling as an allometric cascade, with the b value for overall energy metabolism being determined by the c_i and b_i values for all major steps in the biochemical and physiological pathways of energy demand and energy supply that together make up whole-body MR. Application of this model to evaluate the scaling behaviour of MMR versus BMR illustrates its usefulness.

Our understanding of the control contribution of various steps in ATP demand and supply pathways under MMR conditions (often estimated as maximum oxygen consumption, $\dot{V}O_2\text{max}$) is better than our understanding of control of BMR. This is because the conditions of MMR—by definition, attained when further increments in work intensity do not correspond to further increments in $\dot{V}O_2$ —are readily achievable and reproducible in numerous species. Recent careful statistical and physiological analyses¹⁰ for a group of 'standardized' mammals ($\dot{V}O_2\text{max}$ data standardized to a haemoglobin concentration equal to 15 g per 100 ml and heart mass equal to 1% of body mass) found a scaling coefficient equal to 0.879 (± 0.02 , 95% confidence limits); this b value is highly significantly different from the b value of 0.729 (± 0.038) found for the BMR of the same group of species. This difference¹¹ between BMR and MMR scaling cannot be explained by single-cause scaling theories. To see if our allometric model can account for this difference, we need to analyse the b_i and c_i values for key controlling steps in energy turnover in both conditions. The linked steps on the ATP supply side^{3,4,6} are ventilation, pulmonary diffusion, cardiac output and circulation and capillary-tissue diffusion, plus fuel transfer to cellular catabolic processes in muscles (including fuel mobilization and O_2 transfer to, and consumption by, mitochondria). The b_i coefficients for most of these steps are known and range from ~ 0.7 to ~ 1.0 , and the best available evidence indicates that at MMR control is distributed throughout the energy delivery chain (c_i values ranging from ~ 0.1 to ~ 0.3 (see Methods)).

For energy demand, the situation under MMR conditions is somewhat simplified since over 90% of O_2 consumption^{2,11} is directed towards ATP synthesis for muscle work; that is, for actomyosin ATPase and the Ca^{2+} pump. Scaling studies have usually focused mainly on energy-supply processes, which at MMR may display moderate to zero reserve capacity. In fact, the functional capacities of O_2 delivery systems can sometimes be surpassed. For example, during intense exercise in some human athletes, the functional capacity of the lung at $\dot{V}O_2\text{max}$ is breached and a condition termed 'desaturation' is observed, (blood passing through the lung cannot become fully saturated²); yet, because the short-fall can be made up by adjusting cardiac output and O_2 extraction, $\dot{V}O_2$ can still increase beyond this point. A similar situation can develop for the heart and circulation, which is also resolved by adjusting the sharing of control. (The maximum

metabolic rates per gram of a small muscle mass can increase 2–3-fold above mass-specific rates under whole-body MMR conditions because a larger fraction of cardiac output can be directed to a smaller O_2 -demand site². This means that at whole-body MMR, capillary-mitochondria diffusion and mitochondrial metabolism per se retain reserve capacity.) Despite these adjustments, there is no doubt that at maximum aerobic exercise, O_2 delivery by the lung and heart must be close to an 'upper ceiling'. In contrast, under MMR conditions, actomyosin and Ca^{2+} pump must still display a huge reserve capacity; this is because of the requirement for over 3-fold higher² anaerobic than aerobic ATP turnover rates, primed by the glycolytic path or by phosphocreatine mobilization. Because of these contrasting conditions in the energy-supply versus energy-demand processes as aerobic maximum fluxes are approached, it is not surprising that the control contributions for energy supply increase (as in sum they approach classical rate-limiting function), while those for energy-demand processes under aerobic MMR conditions diminish towards zero^{2,4}. That is why in Fig. 1, to see how they affect scaling we use two c_i values, 0.63 versus 0 and 0.28 versus 0, for actomyosin and Ca^{2+} cycling, respectively.

When the b_i and c_i values for all the above processes are used in equation (2) to see how they contribute to global b values for MMR, we calculate b values that range between 0.92 (b_{max}) to 0.82 (b_{min}), as the c_i values for actomyosin and Ca^{2+} ATPases shift from zero to maximum estimates (Fig. 1). Global b values in the ~ 0.8 to ~ 0.9 range are similar to empirically observed b scaling coefficients¹¹ as well as for MMR data that are standardized with respect to haemoglobin concentration and heart size¹⁰. Thus our model accommodates $\dot{V}O_2\text{max}$ data that show high scaling coefficients for MMR compared to BMR. However, this remains a difficulty in allometric models^{7,8} that assume that constraint on a single function—oxygen delivery—enforces the major b value on metabolism and on all other physiological rate processes. Instead of a physiological delivery system serving as a single rate-limiting process, as required by such models^{7,8}, *in vivo* the empirically observable operational rule of thumb seems to involve sharing of control, with regulatory contributions arising at many key steps in carbon flux, in O_2 flux and in ATP turnover^{3,6}.

Application of the multi-site model to BMR scaling is more difficult because less is known about the control of BMR than of MMR. To identify a starting point, we need to review what processes add up to the whole-body metabolism we refer to as BMR. In the laboratory rat, $\sim 70\%$ of basal whole-body O_2 consumption is used by the mitochondria to produce ATP for several ATP sinks, including protein turnover, Na^+ pump, Ca^{2+} pump, myosin ATPases, gluconeogenesis, ureagenesis, messenger RNA synthesis, and substrate cycling, contributing approximately 20–25, 20–25, 5, 5, 7, 2.5, <2 , and $<2\%$, respectively¹², to BMR.

For energy supply under BMR conditions all the oxygen-delivery steps display a huge excess capacity; this reserve capacity is what allows a 10-fold or more metabolic scope for activity in most animals and is a major reason why (as physiologists have long realised) these steps have minimal effects on the control of metabolism under BMR conditions^{2,10}. Thus, for simulation purposes the c_i values for all the oxygen-delivery steps could be assumed to be zero. If any step were omitted, metabolic fluxes would fall to zero, so it is clear that some control contribution must arise from each of the oxygen-delivery steps, despite their huge excess capacity for oxygen delivery. That is why, in Fig. 1, small but finite c_i values are selected for each of the O_2 delivery steps; if these were set at zero, our overall conclusions, however, would not change.

It should be obvious, given such low c_i estimates for energy supply, that the b_i and c_i values for these steps will contribute little to the global b exponent for BMR (Fig. 1). To find the proximate determinants of the global b value for BMR we must examine the energy demand processes.

For energy demand in the ATP turnover cycle, under BMR

conditions the two major energy sinks, as already mentioned, are the $\text{Na}^+ \text{K}^+$ ATPase and protein turnover^{12,13}, for which b_i values of 0.72 and 0.77, respectively, are available in the literature (see Methods). For $\text{Na}^+ \text{K}^+$ ATPase, c_i values for various tissues vary, so, as explained in the Methods, we use two values, 0.70 and 0.35, to see the effect on the global b value. For protein synthesis the two corresponding c_i values are 0.10 and 0.30, values that bracket the percentage contribution to overall BMR. Finally, for completion we estimate c_i values for the Ca^{2+} pump, for ureagenesis, and for gluconeogenesis; these too are selected to be similar to their percentage contribution to overall BMR.

Using such estimated b_i and c_i values for energy supply and energy demand processes in equation (2), the global b for BMR equals 0.76 when the $\text{Na}^+ \text{K}^+$ ATPase control coefficient is 0.70, and this global b value increases to 0.79 when a control coefficient of 0.35 is associated with this step (Fig. 1b). Our estimates for $\text{Na}^+ \text{K}^+$ ATPase c_i values are conservative and values closer to unity may be appropriate¹³, in which case the global b values for BMR would be even lower. These ranges of global b coefficients are in accord with the well-known 3/4 power law, but are generated largely by b_i and c_i values for ATP-demand processes. They are not generated by constraints imposed by O_2 supply systems, or any other single factor, in contrast to assumptions of most single-cause allometry models. As far as we know, aside from our multiple-causes model, no other allometric models have been able to explain these differing BMR versus MMR scaling patterns. One earlier series of studies¹¹

did emphasize the importance to allometry of co-adaptation of multiple steps in the path of O_2 from air to mitochondria. When viewed from a scaling perspective, we consider that the Weibel–Taylor model (of the matching of several functional links in O_2 flux control¹¹) forms a kind of conceptual base for our currently proposed multi-site model of allometry.

Like $\text{Na}^+ \text{K}^+$ ATPase, the specific activities of many components in the complex pathways of energy supply and energy demand display b_i scaling coefficients close to 0.75, but many do not. For example, in mammals, citrate synthase scales with the value of b_i equal to about 0.9; several other enzymes show their own unique scaling b_i coefficients, and of course their own c_i control coefficients². When coupled or integrated together into a unitary functional system, one could well expect that the b_i and c_i values for the dominant control sites could add up to global b values for metabolism that varied significantly from 0.75. The recent exhaustive analyses of BMR in 487 mammal species, arranged in zoogeographical clusters, show that this variation in b values for whole-body BMR is indeed observed¹⁴. Many of the factors contributing to this variation are referred to in our model as ‘intrinsic’ or ‘extrinsic’ to the proximate metabolic processes accounting for specific allometric behaviours.

Intrinsic factors influencing scaling can be found at various levels of the functional units. For example, not only is the organ-mass scaling pattern organ-specific, but the respiration rate of tissue slices of various organs also follow different, tissue-specific scaling expo-

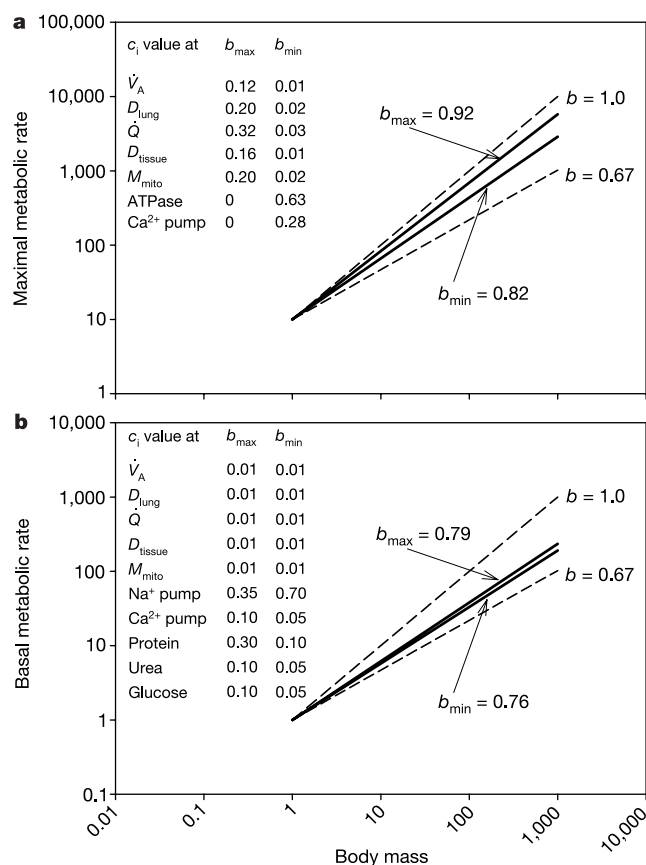


Figure 1 Estimates of b values for maximum and basal metabolic rates in mammals using the multi-site allometry model. The scaling patterns of the maximum and basal metabolic rates (MMR and BMR) were calculated from simulations solving equation (2) using c_i values shown in **a** and **b** (see Methods for b_i values). **a**, For MMR the functions listed in order are alveolar ventilation ($\dot{V}_A$), pulmonary diffusion (D_{lung}), cardiac output ($\dot{Q}$), capillary-mitochondria tissue diffusion (D_{tissue}), cytosolic and mitochondrial metabolism (M_{mito}) plus actomyosin ATPase (ATPase) and the Ca^{2+} pump. The same abbreviations

are used in **b**, in addition to the Na^+ pump and protein, urea and glucose synthesis. The two sets of c_i values yield estimates of maximum ($b_{\max}$) and minimum ($b_{\min}$) values for scaling of MMR and BMR. For simulation purposes, metabolic rates and body mass are given in arbitrary units and the intercept for MMR (**a**) is assumed to be 10-fold higher than for BMR^{10,11} (arbitrarily set at 1). For reference, slopes of 1.0 and 0.67 are shown in **a** and **b**.

nents¹³. These studies confirm differential scaling of various organs and their mitochondrial membrane content, both of which undoubtedly contribute to the scaling of the metabolic rate.

Variation in membrane composition is another intrinsic factor that may influence scaling coefficients. In mammals at rest, 10% of the oxygen consumed by the animal is non-mitochondrial, while another 20% of the O₂ consumed is used to maintain mitochondrial membrane potential affected by the leak of protons¹². Allometric scaling of proton leak across the mitochondrial inner-membrane displays a b_i exponent of 0.86 in mammals¹⁵. This b_i estimate may not be statistically different from 0.75. However, at least tentatively, we assume that the trend is valid, since proton leakiness can be influenced by membrane composition¹⁶, and percentage saturation of biological membranes increases¹³ as body size increases. Although more work is required, it is already evident that membrane composition provides another example of intrinsic effects on metabolic rate, and on scaling not only of the proton leak, but also of the scaling of transport proteins such as Na⁺ K⁺ ATPase¹³.

Although the same, coupled steps usually contribute to the metabolic characteristics of tissues and organs, their quantitative contributions to control may vary in different states. For example, insulin (and presumably endocrines generally) and the kinds of fuels used are intrinsic factors that influence flux rates through specific pathways (such as glycolysis or fatty acid oxidation) and modify control contributions of different sites in ATP turnover².

We mentioned above that the intercept may be viewed as being formed from two ($a_{i,max}$ and f_i) parameters. This subdivision is important when looking across taxa, where current data show that both $a_{i,max}$ and f_i parameters also can be modified. As an example, Suarez *et al.*¹⁷ observed that when comparing among groups (from relatively sluggish fish to super-active bees), each with varying glycolytic flux rates, not only the maximal catalytic capacity of several enzyme sites in glycolysis increased, but at MMR there was also an increase in f_i or fractional velocity. Similarly, when comparing distant groups like ectotherms and endotherms, similar b exponents are observed but an obvious difference in intercept is found. This difference in baseline is in part due to differences in the cost associated with Na⁺ ion pumping for which $a_{i,max}$ is higher in endotherms but f_i is conserved¹³.

Field biologists frequently emphasize that factors such as temperature, pressure, nutritional preferences, water availability and other ecological factors may be 'ultimate causes' or 'ultimate explanations' for why specific groups of organisms show specific allometric behaviour. Appeal, for example, is made to this level of explanation for different scaling exponents for field metabolic rates when comparing reptiles, birds and mammals¹⁸. Such 'ultimate' explanations also apply to the phylogenetically or environmentally linked scaling patterns noted by Lovegrove¹⁴. We interpret all such forces (in addition to purely abiotic ones such as temperature, hydrostatic pressure or osmotic pressure) as extrinsic factors that affect scaling relations by acting on a , b_i , and/or c_i in equation (2) above. So-called 'ultimate causes' thus are not 'causes' in a mechanistic sense, but only in the sense that they act through 'proximal causes' or 'proximal mechanisms' such as those outlined in our model.

In summary, then, most earlier analyses of scaling of metabolic and physiological processes have looked for a single process which scales with a universal b exponent; their models then assumed that this single process enforces the scaling behaviour for all other biochemical and physiological rate functions. We here refer to all such interpretations as single-cause models of scaling. The only way these could be valid is if the selected single process behaved as a master rate-determining step in metabolism and physiology, a concept that has long been abandoned in metabolic regulation research. Taken literally, if a single process, such as O₂ or fuel delivery, is rate-limiting for BMR, there can be no room left for scope for activity: higher metabolic rates would be blocked by this

one process because it would already be rate-limiting even at rest. For this reason and many others², the concept of a master rate-limiting step was replaced by concepts of multi-site controls, with many processes contributing in varying degree to regulation of overall whole-organism metabolic and physiological rates²⁻⁶. When applied to the problem of scaling, the multi-site concept of allometry states that two key parameters (b_i , the scaling exponent and c_i , the control coefficient in equation (2)) for all major control sites in ATP-turnover pathways determine the overall scaling behaviour of whole-organism bioenergetics. For mammals, including humans, these values can be estimated for many of the key steps in energy turnover under two states, BMR and MMR. Our model supplies a mechanistic basis for why the scaling patterns of BMR and MMR differ: in scaling of BMR, the O₂ delivery steps contribute almost nothing to the global b scaling exponent, which is therefore largely determined by energy-demand processes, whereas at MMR, the O₂ delivery steps, now coupled to different ATPases, significantly increase the global b scaling coefficient. □

Methods

Estimating b_i and c_i values for MMR

For energy supply, the b_i and c_i values are estimated from data⁹ for human $\dot{V}O_{2max}$ performance. In these analyses, control contributions from cell metabolism were ignored. However, we assume that mitochondrial metabolism displays a c_i value well above zero (for example, training increases mitochondrial volume by ~25% which correlates with a ~5% increase in $\dot{V}O_{2max}$ ^{2,3}, indicating a c_i value of ~0.2). To accommodate the mitochondrial data and to have all the c_i estimates add up to 1.0 in exercising humans, the estimates from ref. 9 must be multiplied by 1/1.2 or 0.8. Thus for alveolar ventilation $b_i = 0.80$ (ref. 19) and $c_i = 0.12$ (ref. 6; or 0.16 from ref. 9 $\times$ 0.8). For pulmonary diffusion, $b_i = 1.08$ (ref. 11) and $c_i = 0.20$ (ref. 9). For cardiac output²⁰, $b_i = 0.879$ while $c_i = \sim 0.32$ (ref. 9). Capillary volume densities scale with $b_i = 0.89$ (ref. 11), with $c_i = 0.20$ (ref. 9) for O₂ diffusion from capillaries to mitochondria. Mitochondrial volume densities and hence their maximum oxygen consumption rates scale with $b_i = 0.87$ (ref. 11), whereas $c_i = \sim 0.16$ (refs 2, 3).

For energy demand, under MMR conditions over 90% of O₂ consumption^{2,11} is directed towards ATP synthesis for muscle work; that is, for actomyosin ATPase and the Ca²⁺ pump. What we need to complete our calculations are b_i and c_i values for both of these major ATP sinks. For actomyosin ATPase, $b_i = 0.77$ (ref. 21), and at low-to-moderate exercise intensities it contributes significantly to the control of overall metabolic rates³⁻⁵. Allometric data on the Ca²⁺ pump are also not abundant, but because muscle contraction frequency is directly correlated with this activity, an *in vivo* estimate of $b_i = 0.86$ ($r = 0.99$) is provided by scaling of stride frequency in a mouse-to-horse range of body sizes²². Similar scaling coefficients are obtained for the Ca²⁺ pump catalytic capacities in hearts of small and large animals²³. Estimates of c_i values for actomyosin and Ca²⁺ ATPase can be obtained by comparing the impact of change in their capacities upon tissue respiration. Actomyosin concentrations and activities are 3-4-fold higher in FOG (fast twitch oxidative glycolytic) fibres than in slow fibres²⁴, while Ca²⁺ pump densities are about 1.7-fold higher²⁵; together these can 'drive' O₂ delivery about 2.4 times faster^{25,26} (linear over a broad range of submaximal work rates). Actomyosin accounts for a 1.92-fold increase in activation, and Ca²⁺ cycling accounts for the rest (~48% of the activation). Thus under these conditions, the c_i value (change in overall metabolic flux achievable/change in actomyosin catalytic capacity) equals 1.92/3.0 or 0.63, while for the Ca²⁺ pump the c_i equals $\sim 0.48/1.7$ or ~ 0.28 . These values are calculated with respect to O₂ flux, so in total the two processes contribute up to 91% of the control under submaximal work. However, because c_i values may approach zero at MMR¹, in our simulations, we used two c_i values for actomyosin and Ca²⁺ ATPases, 0.63 versus 0 and 0.28 versus 0, respectively. A summary of these values is given in Fig. 1a.

Estimating b_i and c_i values for BMR

For energy supply under BMR conditions all the oxygen-delivery steps display a huge excess capacity¹⁰. Thus, small but finite c_i values were selected for each of these steps. The b_i and c_i values chosen for alveolar ventilation¹⁹ are $b_i = 0.80$ and $c_i = 0.01$. For pulmonary diffusion^{6,11}, $b_i = 1.08$ and $c_i = 0.01$. For cardiac output²⁰ $b_i = 0.76$ and $c_i = 0.01$. For diffusion from capillaries to mitochondria, $b_i = 0.89$ (as above), whereas $c_i = 0.01$. For mitochondrial O₂ consumption, we were confronted with the issue of what state the mitochondria operate in under BMR conditions. As indicated above, state-3 mitochondrial respiration (saturated with all required substrates) is expected to scale to the 0.87 power. However, a positive scaling has been found¹⁶ for the respiratory control ratio (RCR), with a b_i value of 0.121, indicating that state-4 respiration (which may be a probable state for many mitochondria under BMR conditions) scales with a lower exponent than state 3. Thus we considered that b_i values of 0.87 - 0.12 = 0.75 and c_i values of about 0.01 would be reasonable estimates for mitochondrial respiration under BMR conditions.

For the energy demand of the ATP turnover cycle, under BMR conditions the two major energy sinks are the Na⁺ K⁺ ATPase and protein turnover^{2,13}. For Na⁺ K⁺ ATPase, studies¹³ determined an *in vivo* $b_i = 0.72$ and, in our simulation, the c_i was chosen to vary from 0.70 to 0.35. Estimates of c_i values for Na⁺ K⁺ ATPase can be obtained by comparing

the impact of change in $\text{Na}^+ \text{K}^+$ ATPase capacity upon tissue respiration. In vertebrate liver, a 5.6-fold change in $\text{Na}^+ \text{K}^+$ ATPase capacity 'drives' a 1.63-fold change in $\dot{V}\text{O}_2$, while in vertebrate brain, a 3.78-fold change in $\text{Na}^+ \text{K}^+$ ATPase 'drives' a 1.63-fold change in $\dot{V}\text{O}_2$; thus the ratios of change in metabolic flux capacities/change in $\text{Na}^+ \text{K}^+$ ATPase capacities, the c_i values, are 0.29 (for liver) and 0.43 (for brain). We do not have global c_i values for whole-body rates, but because $\text{Na}^+ \text{K}^+$ ATPase activities account for a large percentage of BMR in numerous animals^{12,13}, the c_i values estimated for the liver and brain are probably fairly representative. In some tissues, such as kidneys, the values may be higher; in some, such as non-working muscles, they may be lower¹³. That is why, in our simulation (Fig. 1), two widely differing c_i values, 0.70 and 0.35, are used to illustrate the effect on overall scaling behaviour.

A second major energy-demand process under BMR conditions is protein synthesis. Previous allometric studies of protein synthesis rates establish $b_i = 0.77$ (refs 27, 28), but experimental studies allowing quantifying c_i coefficients for this process are not abundant. Nevertheless, *in vitro* studies²⁹ provide a c_i value of 0.11 for protein synthesis in liver slices of amphibians. In humans and rats, whole-organism studies of protein synthesis and metabolism usually find c_i values that are somewhat higher². For our simulation, with the $\text{Na}^+ \text{K}^+$ ATPase c_i set at 0.70, we selected for protein synthesis a value of $c_i = 0.10$. To complete our simulation, we also included data for three more ATP demand processes: Ca^{2+} ATPase plus urea and glucose biosyntheses. As above, a value of $b_i = 0.86$ ($r = 0.99$) (ref. 22) is our best current estimate for the Ca^{2+} pump, while because of a low contribution to BMR we assume a relatively low value of $c_i = 0.05$. For urea synthesis²⁸ $b_i = 0.77$ and we selected $c_i = 0.05$. Similarly, for gluconeogenesis³⁰, $b_i = 0.76$ and we selected $c_i = 0.05$; the latter two c_i values are again similar to the percentage contributions of these processes to BMR¹². We consider these low estimates to be reasonable, because so far, no studies have found large contributions of ureagenesis or gluconeogenesis to the control of basal metabolism^{3-5,12}. With the $\text{Na}^+ \text{K}^+$ ATPase c_i value set at 0.7, the values for all the ATP demand processes are then 0.10, 0.05, 0.05 and 0.05 respectively; these double when the $\text{Na}^+ \text{K}^+$ ATPase c_i value is reduced to 0.30 (Fig. 1b). The two c_i values for protein synthesis bracket the percentage contribution¹² of this ATP demand process to BMR. A summary of these values is given in Fig. 1b.

Received 5 September 2001; accepted 20 February 2002.

- Kleiber, M. Body size and metabolism. *Hilgardia* **6**, 315–353 (1932).
- Hochachka, P. W. & Somero, G. N. *Biochemical Adaptation—Mechanism and Process in Physiological Evolution* (Oxford Univ. Press, New York, 2002).
- Hochachka, P. W. *Muscles as Molecular and Metabolic Machines* (CRC Press, Boca Raton, Florida, 1994).
- Jeneson, J. A., Westerhoff, H. V. & Kushmerick, M. J. A metabolic control analysis of kinetic controls in ATP free energy metabolism in contracting skeletal muscle. *Am. J. Physiol. Cell Physiol.* **279**, C813–C832 (2000).
- Thomas, S. & Fell, D. A. A control analysis exploration of the role of ATP utilisation in glycolytic-flux control and glycolytic-metabolite-concentration regulation. *Eur. J. Biochem.* **258**, 956–967 (1998).
- Jones, J. H. Optimization of the mammalian respiratory system: symmorphosis versus single species adaptation. *Comp. Biochem. Physiol. B* **120**, 125–138 (1998).
- West, G. B., Brown, J. H. & Enquist, B. J. The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science* **284**, 1677–1679 (1999).
- Banavar, J. R., Maritan, A. & Rinaldo, A. Size and form in efficient transportation networks. *Nature* **399**, 130–132 (1999).
- Wagner, P. D. Algebraic analysis of the determinants of $\dot{V}\text{O}_2$ max. *Resp. Physiol.* **93**, 221–237 (1993).
- Bishop, C. M. The maximum oxygen consumption and aerobic scope of birds and mammals: getting to the heart of the matter. *Proc. R. Soc. Lond. B* **266**, 2275–2281 (1999).
- Weibel, E. R. *Symmorphosis, on Form and Function Shaping Life* (Harvard Univ. Press, Cambridge, 2000).
- Rolfe, D. F. S. & Brown, G. C. Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiol. Rev.* **77**, 731–758 (1997).
- Hulbert, A. J. & Else, P. L. Mechanisms underlying the cost of living in animals. *Annu. Rev. Physiol.* **62**, 207–235 (2000).
- Lovegrove, B. G. The zoogeography of mammalian basal metabolic rate. *Am. Nat.* **156**, 201–219 (2000).
- Porter, R. K. & Brand, M. D. Body mass dependence of H^+ leak in mitochondria and its relevance to metabolic rate. *Nature* **362**, 628–630 (1993).
- Porter, R. K., Hulbert, A. J. & Brand, M. D. Allometry of mitochondrial proton leak: influence of membrane surface area and fatty acid composition. *Am. J. Physiol.* **271**, R1550–R1560 (1996).
- Suarez, R. K., Staples, J. F., Lighton, J. R. B. & West, T. G. Relationship between enzymatic flux capacities and metabolic flux rates: Non-equilibrium reactions in muscle glycolysis. *Proc. Natl Acad. Sci. USA* **94**, 7065–7069 (1997).
- Nagy, K. A., Girard, I. A. & Brown, T. K. Energetics of free-ranging mammals, reptiles, and birds. *Annu. Rev. Nutr.* **19**, 247–277 (1999).
- Stahl, W. R. Scaling of respiratory variables in mammals. *J. Appl. Physiol.* **22**, 453–460 (1967).
- Bishop, C. M. Heart mass and the maximum cardiac output of birds and mammals: implications for estimating the maximum aerobic power input of flying animals. *Phil. Trans. R. Soc. Lond. B* **352**, 447–456 (1997).
- Lindstedt, S. L., Hoppeler, H., Bard, K. M. & Thomson, H. A. Jr Estimates of muscle-shortening rate during locomotion. *Am. J. Physiol.* **249**, R669–R703 (1985).
- Heglund, N. C., Taylor, C. R. & McMahon, T. A. Scaling stride frequency and gait to animal size: mice to horses. *Science* **186**, 1112–1113 (1974).
- Hamilton, N. & Ianuzzo, C. D. Contractile and calcium regulating capacities of myocardia of different sized mammals scale with resting heart rate. *Mol. Cell. Biochem.* **106**, 133–141 (1991).
- Szentesi, P., Zaremba, R., van Mechelen, G. J. M. & Stienen, G. J. M. ATP utilisation for calcium uptake and force production in different types of human skeletal muscle fibres. *J. Physiol. Lond.* **531**, 393–403 (2001).

- Armstrong, R. B. & Laughlin, M. H. Metabolic indicators of fibre recruitment in mammalian muscles during locomotion. *J. Exp. Biol.* **115**, 201–213 (1985).
- Hochachka, P. W., Bianconcini, M., Parkhouse, W. D. & Dobson, G. P. On the role of actomyosin ATPase in regulation of ATP turnover rates during intense exercise. *Proc. Natl Acad. Sci. USA* **88**, 5764–5768 (1991).
- Waterlow, J. C. Protein turnover with special reference to man. *Q. J. Exp. Physiol.* **69**, 409–438 (1984).
- Brody, S. *Bioenergetics and Growth* (Hafner, New York, 1945).
- Fuery, C. J., Withers, P. C. & Guppy, M. Protein synthesis in the liver of *Bufo marinus*—cost and contribution to oxygen consumption. *Comp. Biochem. Physiol. A* **119**, 459–467 (1998).
- Weber, J. M., Fournier, R. & Grant, C. Glucose kinetics of the Virginia possum: Possible implications for predicting glucose turnover in mammals. *Comp. Biochem. Physiol.* **118**, 713–719 (1997).

Acknowledgements

P.W.H. and R.D.A. were supported by the NSERC, Canada; R.K.S. by the NSF in the USA. C.-A.D. was an NSERC and FCAR pre-doctoral fellow.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.W.H. (e-mail: pwh@zoology.ubc.ca).

Host plants influence parasitism of forest caterpillars

J. T. Lill*, R. J. Marquis & R. E. Ricklefs

Department of Biology, University of Missouri—St. Louis, St. Louis, Missouri 63121, USA

Patterns of association between herbivores and host plants have been thought to reflect the quality of plants as food resources^{1,2} as influenced by plant nutrient composition³, defences^{4,5}, and phenology⁶. Host-plant-specific enemies, that is, the third trophic level, might also influence the distribution of herbivores across plant species^{7–10}. However, studies of the evolution of herbivore host range^{11–15} have generally not examined the third trophic level, leaving unclear the importance of this factor in the evolution of plant–insect herbivore interactions¹⁶. Analysis of parasitoid rearings by the Canadian Forest Insect Survey shows that parasitism of particular Lepidoptera species is strongly host-plant-dependent, that the pattern of host-plant dependence varies among species of caterpillars, and that some parasitoid species are themselves specialized with respect to tree species. Host-plant-dependent parasitism suggests the possibility of top-down influence on host plant use. Differences in parasitism among particular caterpillar–host plant combinations could select for specialization of host plant ranges within caterpillar communities. Such specialization would ultimately promote the species diversification of Lepidoptera in temperate forests with respect to escape from enemies.

Here, we test the hypothesis that a caterpillar's host plant influences the identity of its parasitoid enemies and its probability of being parasitized. Our analysis includes a broad range of forest tree species and their associated caterpillar and parasitoid species in southern Ontario and Quebec, Canada, taking advantage of the extensive data set generated by the Canadian Forest Insect Survey¹⁷ (CFIS). Because parasitoids almost always kill their hosts, evidence of host-plant-specific parasitism indicates that members of the third trophic level can influence the suitability of host plants for herbivore

* Present address: Washington University, Tyson Research Center, PO Box 258, Eureka, Missouri 63025, USA.

species. In so doing, they can potentially select for traits that influence host plant use by herbivores.

We selected 15 focal caterpillar species from six families of Lepidoptera represented by a large number of records (around 21,000 total rearing records and more than 4,600 parasitism events). Parasitism levels varied significantly among these herbivores. Across all host plants, parasitism differed among families of Lepidoptera ($\chi^2_{1,5} = 245.2$; $P < 0.0001$), ranging from 11% in the Pyralidae to 27% in the Gelechiidae, and among species ($\chi^2_{1,14} = 810.0$; $P < 0.0001$), ranging from 7 to 39%. This variation presumably represents the outcome of ecological and evolutionary interactions between host caterpillars and their parasitoids, but host plant characteristics may also mediate these interactions^{18,19}. Because the caterpillar species in our sample share many host plants, we examined how both host plant species and host caterpillar species contributed to variation in rate of parasitism. Our analysis included 17 host plant genera, representing 10 families of plants and including both angiosperms and gymnosperms. To our knowledge, only two other studies^{20,21} have examined host-plant-specific parasitism in taxonomically diverse assemblages of herbivores. Both studies showed significant host plant effects, but were limited to only two host plant species. Rearing of individual caterpillar species were generally low (despite several years of intensive collecting) and consequently, host plant effects on parasitism could only be assessed for caterpillars grouped into families or ecological guilds.

In this study, the host plant strongly influenced the level of parasitism experienced by individual species of herbivores. Across the 15 caterpillar species, parasitism was one-third higher on individuals reared from angiosperms (21%) than from gymnosperms (16%) (Wald $\chi^2_{1,1} = 89.5$; $P < 0.0001$). Parasitism differed among host plant families ($\chi^2_{1,7} = 368.1$; $P < 0.0001$), ranging from 16% in Pinaceae to 35% in Salicaceae, and among host plant genera ($\chi^2_{1,16} = 626.9$; $P < 0.0001$), ranging from 8% in *Pinus* spp. to 48% in *Salix* spp. (Fig. 1). Significant host plant effects were also found when caterpillar species were examined individually (Table 1); these host-plant effects were temporally stable (for example, non-significant year $\times$ host plant interactions) for three of the five species with sufficient data for analysis. Thus, host plant identity had a large influence on parasitism levels for both the entire assemblage and for individual caterpillar species.

To separate the effects of caterpillar species from host plant species on variation in parasitism level, we isolated four completely filled portions of the host plant and herbivore data matrix suitable for two-way contingency analysis (Table 2). We based these statistical tests on four subsets of Lepidoptera species $\times$ host plant genus, including two for gymnosperm-feeders and two for angiosperm-feeders. In each case, we were able to estimate the main effects of host plant and caterpillar species as well as the host caterpillar $\times$ host plant interaction on parasitism level.

Significant herbivore and host plant effects were accompanied by highly significant host plant $\times$ herbivore species interactions in all data subsets (Table 2). This study is the first to demonstrate both

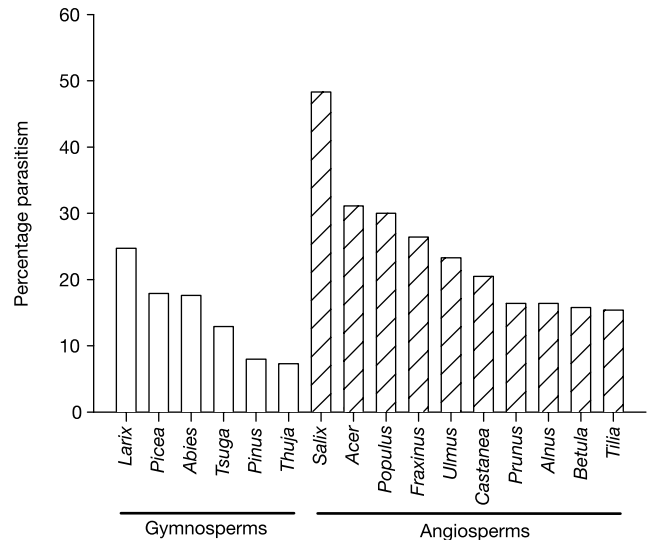


Figure 1 Overall parasitism of 15 focal taxa on 17 host plant genera in two divisions.

that the overall risk of parasitism for a caterpillar assemblage depends upon the identity of the host plant and that the pattern of risk across host plants differs among caterpillar species. Thus, the view that the natural enemies associated with a host plant are a part of its general arsenal of plant defences against herbivores²² requires elaboration: where one caterpillar species is put at risk, another may find refuge.

The host plant–parasitoid matrix is a complex pattern of ‘enemy-reduced’ and ‘enemy-packed’ space (Fig. 2). Here, we designate a host plant as enemy-reduced space for a particular parasitoid species if potential host species of caterpillars reared from that plant have fewer parasitoids than expected. In contrast, potential hosts on enemy-packed plants produced more parasitoids of that particular species than expected. Part of the complex pattern of enemy-reduced and enemy-packed space reflects the heterogeneous nature of the parasitoid community. Our sample yielded 118 species of parasitoids from ten families. Of 39 parasitoid species reared ten or more times, three were restricted to caterpillars found on a single host plant genus and one was a hyperparasitoid; these parasitoids are not included in Fig. 2. The remaining 35 species of parasitoids were reared from caterpillar hosts with broader host plant distributions, allowing us to estimate enemy-reduced and enemy-packed space.

Of the 35 species examined (Fig. 2), most parasitoids exhibited unique, largely non-overlapping patterns of enemy-reduced and enemy-packed space. Six parasitoid species were recovered from individuals of single caterpillar host species collected from a single host plant genus; in each case, individuals of these hosts feeding on other plant genera were free of attack, and in some instances, these

Table 1 Host-plant-specific parasitism levels for six gymnosperm-feeding caterpillar species

Caterpillar species	Host plant genus						All years combined			Host plant $\times$ year	
	<i>Abies</i>	<i>Larix</i>	<i>Picea</i>	<i>Pinus</i>	<i>Thuja</i>	<i>Tsuga</i>	<i>N</i>	χ^2	<i>P</i>	χ^2	<i>P</i>
<i>Acleris variana</i>	0.10	–	0.28	–	–	–	775	36.04	0.0001	4.45	0.49
<i>Caripeta divisata</i>	0.15	0.21	0.12	0.29	–	–	388	5.22	0.1560	–	–
<i>Diorctria abietivorella</i>	0.19	–	0.10	0.13	–	–	1413	5.99	0.0500	1.36	0.82
<i>Lambdina fuscicollaria</i>	0.16	0.13	0.08	0.14	–	0.10	1269	15.49	0.0038	7.78	0.45
<i>Nepytia canosaria</i>	0.14	–	0.08	0.15	0.07	0.15	1236	13.01	0.0112	33.15	0.0001
<i>Semiothisa granitata</i>	0.22	0.30	0.29	0.19	–	0.32	2101	11.19	0.0245	24.73	0.0033

Values are proportion of successfully reared larvae that were parasitized. *N* is the number of total rearings (all years combined) of each species and statistics reported are from categorical analyses. To examine temporal variation in host plant effects, the host plant $\times$ year interaction was also examined for a subset of years with sufficiently large samples (>30 rearings per host plant per year).

Table 2 Main effects and interaction effects of host plant and caterpillar species on the incidence of parasitism

Herbivore species	Host plant genus	χ^2 (herbivore)	χ^2 (plant)	χ^2 (herbivore $\times$ plant)
<i>Lambdina fuscicollis</i> , <i>Semiothisa graminata</i>	<i>Abies</i> , <i>Picea</i> , <i>Pinus</i> , <i>Tsuga</i> , <i>Larix</i>	31.19***	1.39 NS	24.18***
<i>Acleris variaria</i> , <i>Caripeta divisata</i> , <i>Dioryctria abietivorella</i> , <i>L. fuscicollis</i> , <i>Nepytia canosaria</i> , <i>S. graminata</i>	<i>Abies</i> , <i>Picea</i>	108.45***	2.90 NS	74.55***
<i>Archips cerasivorana</i> , <i>H. cuneata</i>	<i>Populus</i> , <i>Prunus</i> , <i>Salix</i>	7.34**	107.63***	29.50***
<i>Choristoneura rosaceana</i> , <i>Hyphantria cunea</i> , <i>Orygia leucostigma</i>	<i>Acer</i> , <i>Betula</i> , <i>Populus</i> , <i>Ulmus</i>	36.27***	37.12***	28.58***

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.0001$; NS, not significant.

plants represented enemy-reduced space. The other parasitoid species were either host caterpillar specialists (10 species in total) reared from a single caterpillar host on multiple host plant genera or generalists (19 species) reared from multiple caterpillar hosts and host plant genera. Ten of 17 host plant genera afforded enemy-reduced space against one or more parasitoids, while 13 showed evidence of enemy packing. *Populus* and *Ulmus* exhibited only enemy-reduced space, while *Acer*, *Fraxinus* and *Pinus* exhibited only enemy-packed space. Of the 267 combinations of parasitoids and host plants, 33 (12.4%) exhibited a significantly reduced rate of parasitism and 29 (10.9%) exhibited a significantly increased rate of

parasitism compared with expected rates based on the relative abundance of potential hosts. Overall, there is a non-clustered distribution of enemy-reduced and enemy-packed space in this matrix, with similar frequencies of each type of space among the three families of parasitoids ($\chi^2 = 4.90$, $P = 0.26$) and the two plant divisions ($\chi^2 = 5.04$, $P = 0.08$). Thus, the suitability of host plants varies with the particular host plant–caterpillar combination. The mechanism(s) contributing to these host plant effects are likely to be varied²¹ and include: (1) plant-volatile-related differences in parasitoid attraction and/or retention rates^{18,23}; (2) variation in the apparency of caterpillars on different host plants²⁴; (3)

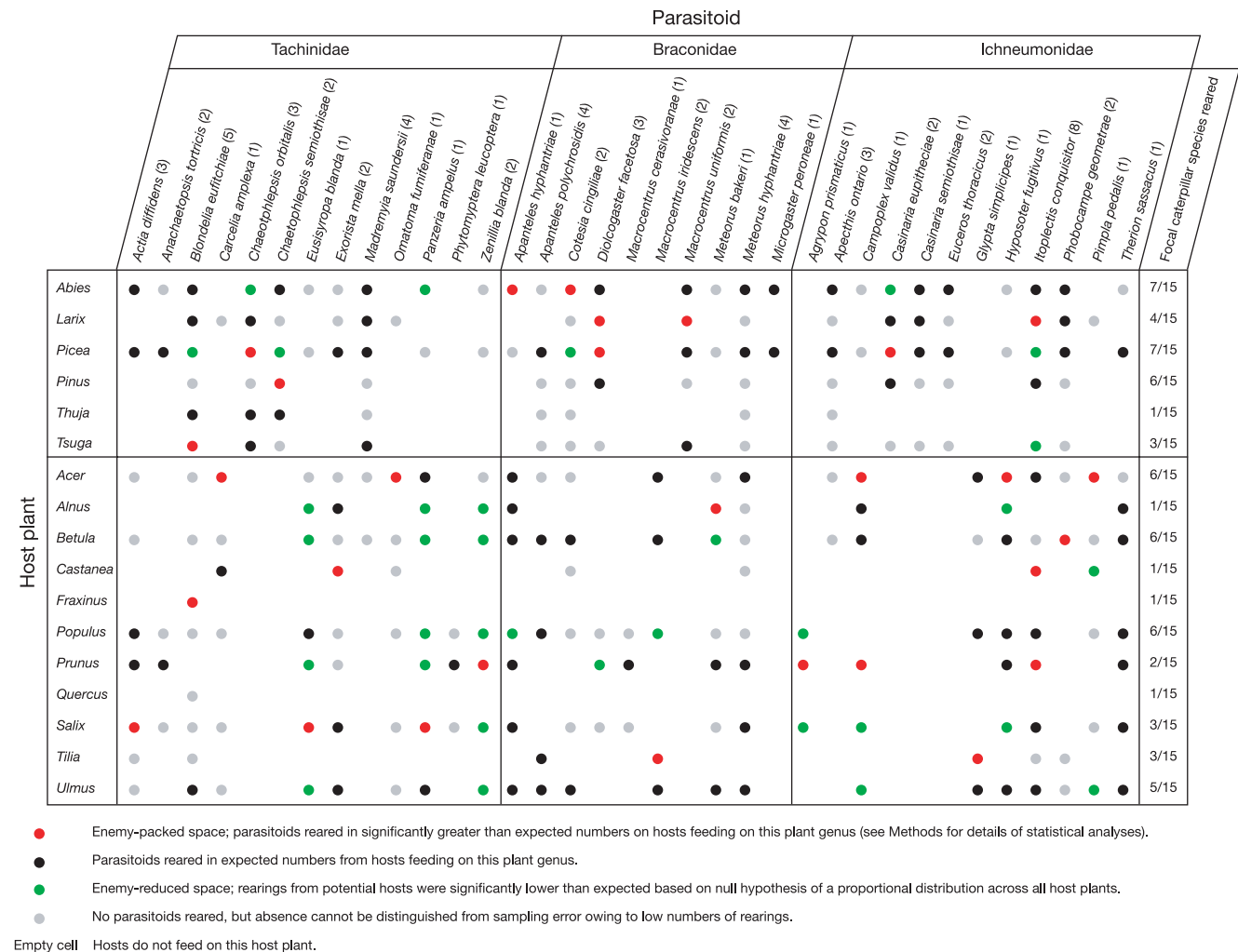


Figure 2 Parasitoid–host plant associations for 35 parasitoid species with hosts that feed on more than one host plant genus. For each parasitoid species, the number of host plants examined reflects the combined diet breadth of the focal caterpillar species that are

potential hosts. The average cell size (± 1 s.e.m.) is 544 ± 64.5 and ranges from 31 to 6,652. The number of focal caterpillar species that are host to each parasitoid is shown in parentheses.

density-dependent foraging by parasitoids (if host abundances also vary with host plant); (4) host-plant effects on caterpillar resistance to parasitism^{25,26}; and (5) the presence and/or abundance of other herbivore species on the host plants (that is, apparent competition via shared natural enemies)^{10,27}. As found previously²¹, there was no relationship in this study between risk of parasitism and the abundance of a caterpillar species on a particular host plant, either for all 15 caterpillar species together ($r^2 = 0.0002$, $N = 62$, $F_{1,60} = 0.012$, $P = 0.91$) or for each species individually (mean $r^2 = 0.10 \pm 0.025$, $N = 5-8$, $P = 0.30-0.92$ for the seven most polyphagous species), suggesting that future research should focus on other possible mechanisms.

These results demonstrate that a caterpillar's host plant, a fundamental aspect of its niche, influences the vulnerability of that insect to its natural enemies. The strong influence of host plant on the risk of attack by parasitoids suggests that the third trophic level should be considered a potentially important selective force in the evolution of herbivore diet breadth. Given a suite of 'suitable' host plants that nevertheless vary in quality with respect to caterpillar growth and development in the absence of enemies, we predict that selection by the third trophic level could favour either increased or decreased host plant specialization, depending on how the risk of parasitism varies with food quality among host plants²⁸. Regardless of that relationship, our data suggest that the third trophic level, in addition to host plant quality, should be considered as a contributing factor influencing the diversification of herbivore faunas. □

Methods

Database

We entered all data yielding parasitoids and adult moths (1937–1947) from hand-written rearing records provided by the CFIS (now the Forest Insect Disease Survey), Sault Ste Marie, Ontario, Canada. Collection records of caterpillars from each tree species and site include the number of larvae of each species collected, the number of adult moths emerging, and the number of parasitoids, mostly identified to species¹⁷. Collections were made from host plants that overlap considerably in distribution²⁹. Although collections contained several individuals per caterpillar–host plant combination, no single collection contained more than 2% of the total rearings for a caterpillar species; thus, we considered all rearings independent in the analyses. We updated the taxonomy of host plants, herbivores and parasitoids when possible, using the current literature. From this larger data set, we selected 15 focal species having moderate to large numbers of records of parasitized individuals. For these, we entered the additional rearing records from which no parasitoids emerged. The focal taxa, each of which belonged to a different genus, span six families of Lepidoptera (Arctiidae [*Hyphantria cunea*], Gelechiidae [*Anacamptis innocuellia*], Geometridae [*Alsephila pometaria*, *Caripeta divisata*, *Erannis tiliaria*, *Lambdina fiscellaria*, *Nepytia canosaria*, *Semiothisa graminata*], Lymantriidae [*Orgyia leucostigma*], Pyralidae [*Dioryctria abietivorella* and *Tortricidae*] [*Acleris variana*, *Archips cerasivorana*, *Choristoneura rosaceana*, *Rhyacionia buoliana* and *Pseudosciaphila duplex*]) and feed on host trees in ten families and seventeen genera, including both angiosperms and gymnosperms. Individual focal taxa varied in diet breadth, from highly restricted (feeding on a single host plant genus) to highly polyphagous (hosts in over 15 plant families).

Data analysis

Data from different sites and years were pooled for analysis. We constructed a $15 \times 17 \times 2$ matrix of the number of parasitized and non-parasitized individuals of each of 15 caterpillar species on each of 17 host plant genera. For a given caterpillar species, we included only host plant genera from which at least 30 individuals (parasitized + non-parasitized) had been reared (cellmean ± 1 s.e.m. = 332.3 ± 96.7). Using nested sets of categorical models (SAS PROC CATMOD) with proportion parasitism as the response variable, we examined the individual effects of host plant division (angiosperm versus gymnosperm), host plant family (including families with over 100 rearing records), host plant genus (including genera containing over 100 rearing records), herbivore family and herbivore species. To examine the joint and interactive effects of herbivore species and host plant genus, we were restricted to analysing subsets of the matrix (see Table 2) because 75% of the cells in the matrix were empty, owing to variation in the host ranges of different herbivores.

Because parasitism also may depend on a species' relative abundance on a particular host, we regressed parasitism on abundance (log-transformed), using the number of collections made on each host plant genus over the entire CFIS collection period (1933–1957) as an estimate of relative abundance. We did this for the combined set of 15 caterpillar species and for seven species that individually fed on five or more host plants. Proportion parasitism was arcsine square-root transformed before analysis to improve normality.

To examine patterns of parasitism of potential hosts (caterpillar species from which at least one individual had been reared) by individual parasitoid species across host plants

(Table 2), we used Monte Carlo $2 \times N$ randomization tests³⁰ (1,000 randomizations per parasitoid species); when parasitism varied significantly among host plants, we performed Fisher's exact tests³⁰ comparing parasitism on each host plant to parasitism on all of the other host plants combined (dividing alpha by the number of tests performed). This is analogous to the procedures for subdividing contingency tables used in standard chi-square tests, where expectations are sufficiently large to permit standard approximations. When the Fisher's exact tests were significant, and depending on the directionality of the effect (greater or less than expected), host plants were labelled as enemy-packed or enemy-reduced, respectively, for a particular parasitoid species.

Received 17 December 2001; accepted 31 January 2002.

1. Ehrlich, P. R. & Raven, P. H. Butterflies and plants: a study of coevolution. *Evolution* **18**, 586–608 (1964).
2. Cates, R. G. Feeding patterns of monophagous, oligophagous, and polyphagous insect herbivores: the effect of resource abundance and plant chemistry. *Oecologia* **46**, 22–31 (1980).
3. Scriber, J. M. & Feeny, P. Growth of herbivorous caterpillars in relation to feeding specialization and to the growth form of their food plants. *Ecology* **60**, 829–850 (1979).
4. Courtney, S. P. Coevolution of pierid butterflies and their cruciferous host plants. III. *Anthracis cardamines* (L.) survival, development, and oviposition on different host plants. *Oecologia* **51**, 91–96 (1981).
5. Becerra, J. X. Insects on plants: macroevolutionary chemical trends in host use. *Science* **276**, 253–256 (1997).
6. Wood, T. K. & Keese, M. C. Host-plant-induced assortative mating in *Enchenopa* tree hoppers. *Evolution* **44**, 619–628 (1990).
7. Brower, L. P. Bird predation and food plant specificity in closely related procrystic insects. *Am. Nat.* **864**, 886–892 (1958).
8. Hawkins, B. A. & Lawton, J. H. Species richness for parasitoids of British phytophagous insects. *Nature* **326**, 788–790 (1987).
9. Bernays, E. & Graham, M. On the evolution of host specificity in phytophagous arthropods. *Ecology* **69**, 886–892 (1988).
10. Holt, R. D. & Lawton, J. H. The ecological consequences of shared natural enemies. *Annu. Rev. Ecol. Syst.* **25**, 495–520 (1994).
11. Feder, J. L., Chilcote, C. A. & Bush, G. L. Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. *Nature* **336**, 61–64 (1988).
12. Gould, F. Rapid host evolution in a population of the phytophagous mite *Tetranychus urticae* Koch. *Evolution* **33**, 791–802 (1979).
13. Bush, G. L. Modes of animal speciation. *Annu. Rev. Ecol. Syst.* **6**, 339–364 (1975).
14. Rausher, M. D. Variability for host preference in insect populations: mechanistic and evolutionary models. *J. Insect Physiol.* **31**, 873–889 (1985).
15. Thompson, J. N. *The Coevolutionary Process* (Univ. Chicago Press, Chicago, 1994).
16. Jaenike, J. Host specialization in phytophagous insects. *Annu. Rev. Ecol. Syst.* **21**, 243–273 (1990).
17. McGugan, B. M. The Canadian Forest Insect Survey. *Proc. 10th Int. Congr. Entomol.* **4**, 219–232 (1958).
18. De Moraes, C. M., Lewis, W. J., Pare, P. W., Alborn, H. T. & Tumlinson, J. H. Herbivore-infested plants selectively attract parasitoids. *Nature* **393**, 570–573 (1998).
19. Barbosa, P. & Benrey, B. in *Conservation Biological Control* (ed. Barbosa, P.) 55–82 (Academic, San Diego, 1998).
20. Le Corff, J., Marquis, R. J. & Whitfield, J. B. Temporal and spatial variation in a parasitoid community associated with the herbivores that feed on Missouri *Quercus*. *Environ. Entomol.* **29**, 181–194 (2000).
21. Barbosa, P. et al. Differential parasitism of macrolepidopteran herbivores on two deciduous tree species. *Ecology* **82**, 698–704 (2001).
22. Janzen, D. H. A host plant is more than its chemistry. *Illinois Nat. Hist. Bull.* **33**, 141–174 (1985).
23. Elzen, G. W., Williams, H. J. & Vinson, S. B. Response by the parasitoid *Campoplex sonorensis* (Hymenoptera: Ichneumonidae) to chemicals (synomones) in plants: implications for host habitat location. *Environ. Entomol.* **12**, 1873–1877 (1983).
24. Weseloh, R. M. in *Caterpillars: Ecological and evolutionary Constraints on Foraging* (eds Stamp, N. E. & Casey, T. M.) 203–221 (Chapman & Hall, New York, 1993).
25. Greenblatt, J. A. & Barbosa, P. Effects of hosts' diet on two pupal parasitoids of the gypsy moth: *Brachymeria intermedia* (Nees) and *Coccygomimus turionellae* (L.). *J. Appl. Ecol.* **18**, 1–19 (1981).
26. Mueller, T. F. The effect of plants on the host relations of a specialist parasitoid of *Heliothis* larvae. *Ent. Exp. Appl.* **34**, 78–84 (1983).
27. Gross, P. & Price, P. W. Plant influences on parasitism of two leaf miners: a test of enemy-free space. *Ecology* **69**, 1506–1516 (1988).
28. Ohsaki, N. & Sato, Y. Food plant choice of *Pieris* butterflies as a trade-off between parasitoid avoidance and quality of plants. *Ecology* **75**, 59–68 (1994).
29. Karban, R. E. & Ricklefs, R. E. Host characteristics, sampling intensity, and species richness of Lepidoptera larvae on broad-leaved trees in southern Ontario. *Ecology* **64**, 636–641 (1983).
30. Engels, W. *Statistical Tests for Categorical Data* Version 1.0 (software package) (Univ. Wisconsin, Madison, 1988).

Acknowledgements

We thank G. Howse for making available the CFIS data, E. Selig, A. Frevert, J. McGrath, and M. Walker for their efforts in constructing the database, B. Hawkins, K. Boege, G. Chen, R. Forkner, S. Renner, R. Rios and K. Schultz for their comments on the manuscript, and P. DeVries for software suggestions. Financial support for this project was provided by a University of Missouri Research Award to R.E.R. and R.J.M. Support for J.T.L. was provided by the Missouri Department of Conservation and the USDA.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.T.L. (e-mail: lillj@biology2.wustl.edu).

Imperfect optics may be the eye's defence against chromatic blur

James S. McLellan*, Susana Marcos*†, Pedro M. Prieto*‡ & Stephen A. Burns*

* Schepens Eye Research Institute, 20 Staniford Street, Boston, Massachusetts 02114, USA

† Instituto de Óptica, Daza de Valdés, Consejo Superior de Investigaciones Científicas, 28006 Madrid, Spain

‡ Laboratorio de Óptica, Universidad de Murcia, Edificio C, Campus Espinardo, 30071 Murcia, Spain

The optics of the eye cause different wavelengths of light to be differentially focused at the retina. This phenomenon is due to longitudinal chromatic aberration, a wavelength-dependent change in refractive power¹. Retinal image quality may consequently vary for the different classes of cone photoreceptors, cells tuned to absorb bands of different wavelengths. For instance, it has been assumed that when the eye is focused for mid-spectral wavelengths near the peak sensitivities of long- (L) and middle- (M) wavelength-sensitive cones, short-wavelength (bluish) light is so blurred that it cannot contribute to and may even impair spatial vision^{2,3}. These optical effects have been proposed to explain the function of the macular pigment⁴, which selectively absorbs short-wavelength light, and the sparsity of short-wavelength-sensitive (S) cones⁵. However, such explanations have ignored the effect of monochromatic wave aberrations present in real eyes. Here we show that, when these effects are taken into account, short wavelengths are not as blurred as previously thought, that the potential image quality for S cones is comparable to that for L and M cones, and that macular pigment has no significant function in improving the retinal image.

In natural, polychromatic light, the retinal image is affected by interactions among longitudinal chromatic aberration (LCA), wave aberrations and transverse chromatic aberration (TCA). Wave aberrations are local deviations in the path of light entering the eye through different points in the pupil. They can be caused by irregularities in the cornea and lens, for example, and they can produce complex distortions in the retinal image even in monochromatic light. TCA is an angular displacement across the retina of the images for different wavelengths. Whereas LCA is dependent

solely on the change in refractive index of the eye's optical elements with wavelength, TCA derives both from this index change and from misalignments between the cornea, lens and fovea^{6,7}. The range of LCA is very consistent across individuals, but wave aberrations and TCA are subject to individual differences^{8,9}. We measured these quantities at a series of wavelengths in three subjects to investigate the image quality of the eye in white light.

Retinal image quality can be characterized by the modulation transfer function (MTF), a spatial-frequency-dependent measure of the relative contrast attenuation from object to image caused by the optics of the eye. The MTF is computed as the modulus of the Fourier transform of the point spread function (PSF), the retinal image produced by an infinitely distant point light source. In Fig. 1 the thin solid curve represents the optimal MTF for a 550-nm monochromatic source, imaged through a 6-mm pupil in a theoretical model eye with LCA but no other aberrations. The thin dotted curve shows the MTF for the same model eye for a 450-nm source that is blurred by 0.7 dioptres. This is the amount of defocus that would result from LCA in the human eye if the eye were optimally focused for 550 nm. At 450 nm, the MTF is degraded, with contrast at some spatial frequencies reduced to zero. The thick curves show the corresponding MTFs for one of the subjects in this study with the empirically determined wave aberrations and the same degree of LCA-induced defocus at 450 nm relative to 550 nm as the model eye. For this subject, the 550-nm (solid) and 450-nm (dotted) curves are very similar. In real eyes, the wave aberrations decrease the variability in MTF across wavelengths that would occur in an eye subject to LCA alone. Figure 2a summarizes this effect for the three subjects, showing the area under the MTF (calculated for 0–100 cycles deg⁻¹) as a function of wavelength. The range of LCA for these subjects was consistent with published values¹. For each subject, the degree of defocus at 550 nm was offset from 0 dioptres to optimize the area under the MTF at this wavelength. This constant offset was then added to the defocus at each wavelength to preserve the range of LCA across the spectrum. In the theoretical model eye with LCA alone (thick curve), MTF area decreases on either side of a sharp peak, varying by 27-fold from minimum to maximum. However, for the three subjects, MTF area is relatively constant across all wavelengths, with a mean variation of less than twofold. The monochromatic aberrations in the real eyes attenuate the MTF at mid-spectral wavelengths, but they actually improve the MTFs at the spectral extremes. This does not imply that image quality at any given wavelength could not be further improved with the appro-

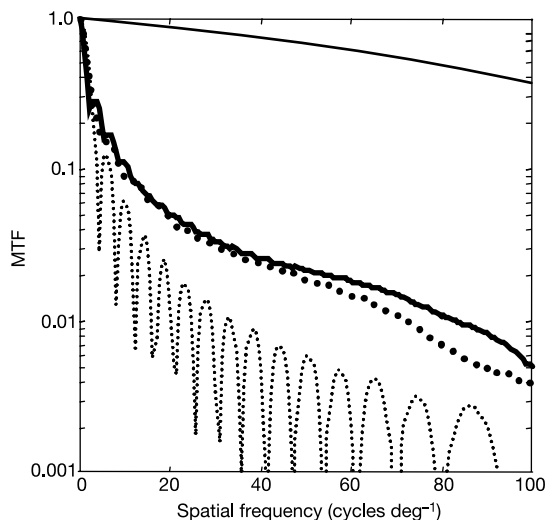


Figure 1 MTFs at 550 nm (solid curves) and 450 nm (dotted curves) for two eyes: (1) a theoretical model eye subject to LCA but with no wave aberrations (thin lines) and (2) a real human subject's eye (thick lines). The other two subjects showed similar results. The

MTFs shown here and in subsequent figures are radial averages of two-dimensional MTFs.

appropriate defocus correction under monochromatic conditions. Figure 2b compares the three subjects' mean MTF area as computed above (solid curve) with the mean MTF area attained when each wavelength's MTF area is individually optimized (dashed curve). This improves both long-wavelength and short-wavelength MTFs, producing a spectrally flat function for these subjects.

Of course, the visual system generally operates under polychromatic conditions, and the small variability in MTFs across the spectrum afforded by the wave aberrations has several consequences for the retinal image available to the cone photoreceptors in natural light. Because wave aberrations are measured under and defined for monochromatic light, the PSF for polychromatic light has been simulated computationally here by compositing the monochromatic PSFs for individual wavelengths. The MTF from this composite PSF gives a measure of polychromatic retinal image quality. The MTF associated with a specific spectral sensitivity function, for example the S-cone fundamental, is computed from a weighted PSF, in which the weights represent the relative sensitivity at each wavelength. The polychromatic PSF must also take into account each individual's TCA, which shifts the position of each monochromatic PSF on the retina. Figure 3a shows the MTFs for the L-cone, M-cone and S-cone fundamentals¹⁰ in equal-energy white light for the model eye with LCA alone and with optimal resolution at 550 nm. In this case, the S-cone MTF lies well below those of L and M cones, because of the image blur caused by LCA at short wavelengths. The individual peaks in the S-cone function represent spatial frequency ranges of alternating contrast reversal, in which light and dark are exchanged. Figure 3b–d shows MTFs computed for each cone class for the eyes of the three subjects, taking into account each subject's LCA, TCA and monochromatic wave aberrations at a series of wavelengths. Their wave-aberration and TCA magnitudes are given in the legend to Fig. 3. The same defocus adjustment as described above to optimize the MTF volume at 550 nm was used. Even when the MTF is thus optimized for mid-spectral light, and despite individual differences in TCA and wave aberrations, the potential image quality for the S cones approaches, and sometimes exceeds, that of L and M cones for all subjects.

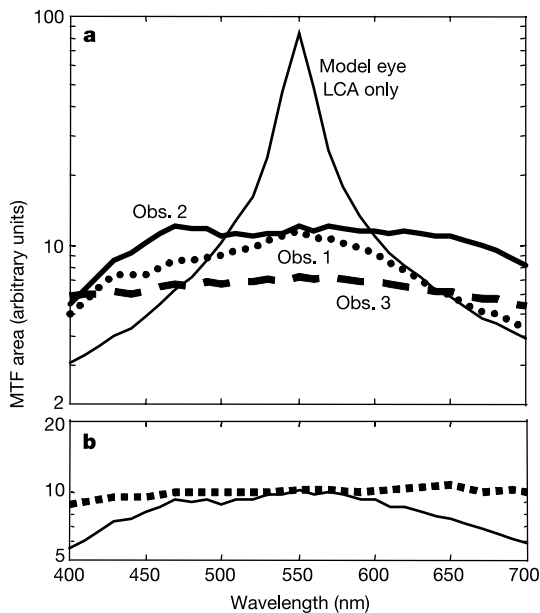


Figure 2 MTF area. **a**, Area under the MTF (arbitrary units) as a function of wavelength for a theoretical model eye with LCA only and for three subjects with measured wave aberrations. **b**, Mean MTF area for all three subjects when defocus is set to optimize area at 550 nm (solid line) and when each wavelength is individually optimized (dashed line). The dashed line shows that MTF area at any single wavelength can be improved further by correcting focus at that wavelength.

Furthermore, regions of contrast reversal are eliminated. These calculations do not take into account the different sampling mosaics for the different cone classes; they simply show the potential image quality for each cone class if the cone classes were to be similarly distributed in the retina. Although the MTF of the S cones is degraded much less by LCA than has been conventionally assumed, the sparsity of S cones still limits their spatial sensitivity. In addition, individuals with exceptionally low wavefront error might have cone MTFs more similar to those shown for the LCA-only theoretical eye. However, the subjects used here have wave aberrations that are representative of the general population¹¹.

The S-cone fundamental's relatively narrow spectral bandwidth compared with those of L and M cones also contributes to potential S-cone image quality by limiting the influence of TCA displacement in the composite PSF. However, because the L and M cones have relatively wide bandwidths, with sensitivity extending into the short-wavelength range, the polychromatic images available to these cones also benefit from the aberration-induced reduction in variability in monochromatic MTFs. Without wave aberrations, foveal vision might be subject to differential blurring of objects with different spectral reflectances and to contrast reversals caused by changes in an object's spatial frequency content with the subject's movement through the environment. Perceptual artefacts could also occur with large changes in the spectral distribution of the environmental illumination.

The improvement in short-wavelength image quality shown here brings into question one possible role of macular pigment in the eye. Macular pigment selectively absorbs wavelengths below 530 nm, with a peak absorption at 460 nm (ref. 12). Although it has often been proposed that macular pigment attenuates short-wavelength light at the retina to alleviate the deleterious effects of LCA, Fig. 4 shows that macular pigment actually provides little if any improvement in the MTF. The two upper curves represent polychromatic MTFs for the LCA-only theoretical eye for two luminance functions: (1) a combination of L-cone and M-cone fundamentals¹⁰ that does not include macular pigment and (2) the same luminance function including macular pigment¹² with a peak absorption of 90%. The

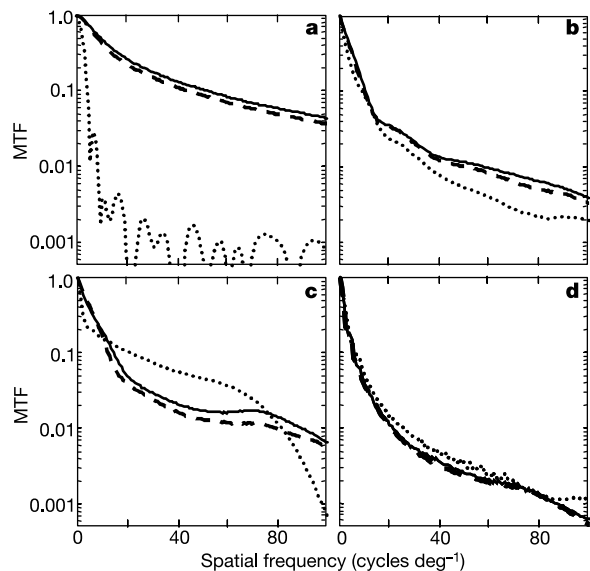


Figure 3 Polychromatic MTFs computed with a 6-mm pupil for L cones (dashed line), M cones (solid line) and S cones (dotted line) for theoretical model eye with LCA only (**a**) and three subjects (**b**, observer 1; **c**, observer 2; **d**, observer 3) with measured LCA, TCA and wave aberrations. RMS wavefront errors (at 530 nm) for the three subjects were 1.83, 0.84 and 2.11 μm . Their TCA magnitudes were: 0.30, 0.13 and 0.18 arcmin nm^{-1} . Similar results were found with a 4-mm pupil (not shown). A comparison of the average results for 6-mm and 4-mm pupils is shown in Supplementary Information.

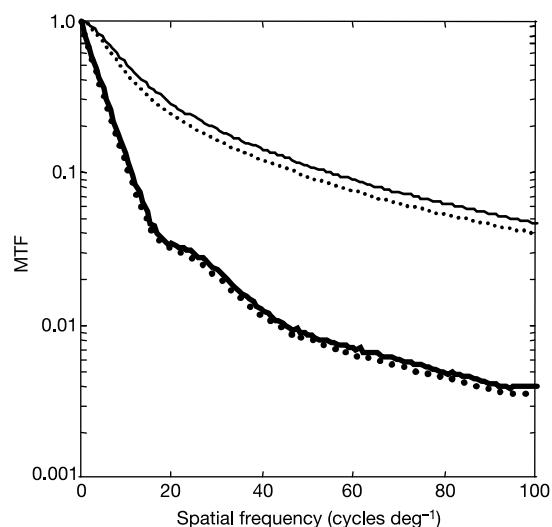


Figure 4 Polychromatic MTFs for the theoretical model eye (thin lines) and for one subject (thick lines) with (solid lines) and without (dotted lines) the effect of macular pigment absorption at short wavelengths. Macular pigment produces only a trivial improvement in the MTF. The other two subjects showed similar results.

two lower curves represent MTFs under the same conditions for one subject in this study. The other subjects showed similar results. For the model eye, macular pigment provides at best a 20% improvement in MTF. For the real eye, the two MTFs are nearly identical. Although macular pigment might serve very important purposes in the eye, such as protecting the retina from oxidative damage from short-wavelength light^{13–15}, it apparently does not affect image quality.

It has been widely assumed that chromatic defocus from the eye's optics degrades the retinal image of short-wavelength light. But this assumption has not previously been tested in a manner that takes into account all of the eye's optical aberrations, measured at multiple wavelengths. We have shown that there is actually little variability in the eye's image quality, as quantified by MTF, across the visible spectrum. Wave aberrations cause the visual system to sacrifice resolution at a single wavelength but allow it to gain approximate constancy in spatial sensitivity across the spectrum. This constancy might provide an even more effective solution to the problems of chromatic blur than could be attained by attenuation and sparse sampling of short-wavelength light in an eye with perfect optics. □

Methods

Apparatus

Wave-aberration data were collected psychophysically with a spatially resolved refractometer¹⁶. For an array of pupil entry locations, the subject visually aligns a monochromatic test spot with a stationary reference location on the retina that enters through the pupil centre. Moving the test spot on the retina corresponds to a change in its angle of incidence at the pupil. The angle needed to align the spot with the reference position for each pupil entry location provides an estimate of the wavefront slope at that location. In a separate optical channel, a fixation cross with a high-spatial-frequency background (namely text) is displayed through a small, centred pupil. The cross is used as both the fixation target and the reference point for aligning the test spot; the text acts as an accommodative cue. A Wratten 58 (green) filter is used in this channel to limit the illumination spectrum of a fluorescent source and to maintain accommodation in the mid-spectral range.

Subjects

Three adult subjects, one female and two male (ages 27, 36 and 48 years), participated in this study. Measurements were performed on the right eye only. Each subject's pupil was dilated with 0.5% tropicamide solution to ensure that all test spots would be visible. The effective diameter of the entire pupil-sampling array is 7.32 mm.

Procedure

The subject's head was stabilized with a dental-impression bite bar on a three-dimensional translating stage. The experimenter monitored the subject's pupil position to align the

pupil centre to the optical axis of the apparatus to correct the position during experimental runs. In each run, the test spot was stepped through 37 entrance pupil locations in pseudo-random order, and the subject used a joystick to move the spot's image to the centre of the fixation cross for each pupil location. Aberrations were measured at six different wavelengths (450, 490, 530, 570, 620 and 650 nm) by placing interference filters (10 nm half-width) in the test spot channel. Each subject completed three runs at each wavelength. Spherical refractive error was subjectively corrected for each subject by means of a translating focusing block to bring the high-frequency background into focus. The same spherical correction was used for all runs at all wavelengths.

Data analysis and computational methods

The data from the spatially resolved refractometer provide estimates of the local slope of the wave aberration at different pupil locations. Zernike polynomial coefficients up to the seventh order (35 Zernike terms) were determined by least-squares fits of the derivatives of the Zernike polynomials to these data. For each subject, each run was fitted individually, and the mean Zernike coefficients across the three runs for each wavelength were used to reconstruct the wave aberration in the pupil plane for that wavelength. Zernike coefficients for wavelengths from 400 to 700 nm were interpolated and extrapolated at 10-nm steps from a polynomial fit to the coefficients for the six measured wavelengths¹⁶. The change in second-order Zernike defocus with wavelength was used as an estimate of LCA. Transverse chromatic aberration (TCA) was determined by measuring wave aberrations at two distinct wavelengths (476 and 601 nm) by using a single magenta filter. TCA was estimated as a linear function of the difference in PSF positions for these two wavelengths and extrapolated from 400 to 700 nm.

The reconstructed wavefronts were used to compute the monochromatic PSF and two-dimensional MTF for each wavelength. A pupil diameter of 6 mm was used for all PSF and MTF calculations. This is near the average pupil size for 20–40-year-olds under illumination on the order of 100 cd m⁻² (ref. 17). One-dimensional MTFs were calculated as the radial average of the two-dimensional MTFs.

Polychromatic PSFs for the cone spectral sensitivity functions¹⁰ and macular pigment¹² were created by summing the PSFs for each wavelength after they had been (1) scaled by the spectral sensitivity for that wavelength and (2) linearly translated to account for TCA at that wavelength. For each subject, second-order defocus was individually adjusted by a constant value to produce optimal MTF volume to 100 cycles deg⁻¹ at 550 nm. This constant was added to the defocus term at all wavelengths to maintain the range of LCA across the spectrum.

Received 23 July 2001; accepted 12 February 2002.

1. Bedford, R. E. & Wyszecki, G. W. Axial chromatic aberration of the human eye. *J. Opt. Soc. Am.* **47**, 564–565 (1957).
2. Boynton, R. M. *Human Color Vision* (Holt, Rinehart & Winston, New York, 1979).
3. Mollon, J. D. A taxonomy of tritanopias. *Docum. Ophthalmol. Proc. Ser.* **33**, 87–101 (1982).
4. Reading, V. M. & Weale, R. A. Macular pigment and chromatic aberration. *J. Opt. Soc. Am.* **64**, 231–234 (1974).
5. Mollon, J. D. "Tho' she kneel'd in that place where they grew..." The uses and origins of primate colour vision. *J. Exp. Biol.* **146**, 21–38 (1989).
6. Marcos, S., Burns, S. A., Prieto, P. M., Navarro, R. & Baraibar, B. Investigating sources of variability of monochromatic and transverse chromatic aberrations across eyes. *Vis. Res.* **41**, 3862–3871 (2001).
7. Thibos, L. N., Ye, M., Zhang, X. & Bradley, A. Spherical aberration of the reduced schematic eye with elliptical refracting surface. *Optom. Vis. Sci.* **74**, 548–556 (1997).
8. Rynders, M., Lidkea, B., Chisholm, W. & Thibos, L. N. Statistical distribution of foveal transverse chromatic aberration, pupil centration, and angle psi in a population of young adult eyes. *J. Opt. Soc. Am.* **12**, 2348–2357 (1995).
9. McLellan, J. S., Marcos, S. & Burns, S. A. Age-related changes in monochromatic wave aberrations of the human eye. *Invest. Ophthalmol. Vis. Sci.* **42**, 1390–1395 (2001).
10. Stockman, A., MacLeod, D. I. A. & Johnson, N. E. Spectral sensitivities of the human cones. *J. Opt. Soc. Am.* **10**, 2491–2521 (1993).
11. Porter, J., Guirao, A., Cox, I. G. & Williams, D. R. Monochromatic aberrations of the human eye in a large population. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **18**, 1793–1803 (2001).
12. Wysecki, G. & Stiles, W. S. *Color Science: Concepts and Methods, Quantitative Data and Formulae* (Wiley, New York, 1982).
13. Haegerstrom-Portnoy, G. Short-wavelength-sensitive-cone sensitivity loss with aging: a protective role for macular pigment? *J. Opt. Soc. Am.* **5**, 2140–2144 (1988).
14. Snodderly, D. M. Evidence for protection against age-related macular degeneration by carotenoids and antioxidant vitamins. *Am. J. Clin. Nutr.* **62** (6 Suppl.), 1448S–1461S (1995).
15. Landrum, J. T. & Bone, R. A. Lutein, zeaxanthin, and the macular pigment. *Arch. Biochem. Biophys.* **385**, 28–40 (2001).
16. Marcos, S., Burns, S. A., Moreno-Barriuso, E. & Navarro, R. A new approach to the study of ocular chromatic aberrations. *Vis. Res.* **39**, 4309–4323 (1999).
17. Winn, B., Whitaker, D., Elliot, D. B. & Phillips, N. J. Factors affecting light-adapted pupil size in normal human subjects. *Invest. Ophthalmol. Vis. Sci.* **35**, 1132–1137 (1994).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

The authors thank D. Williams for comments on the manuscript.

Correspondence and requests for materials should be addressed to J.S.M. (e-mail: mclellan@vision.eri.harvard.edu).

Transient aggregation of ubiquitinated proteins during dendritic cell maturation

Hugues Lelouard*, Evelina Gatti*, Fanny Cappello, Olivia Gresser, Voahirana Camosseto & Philippe Pierre

Centre d'Immunologie de Marseille-Luminy, CNRS-INSERM-Université Med., Campus de Luminy, Case 906, 13288 Marseille, Cedex 09, France

* These authors contributed equally to this work

Dendritic cells (DCs) are antigen-presenting cells with the unique capacity to initiate primary immune responses¹. Dendritic cells have a remarkable pattern of differentiation (maturation) that exhibits highly specific mechanisms to control antigen presentation restricted by major histocompatibility complex (MHC)². MHC class I molecules present to CD8⁺ cytotoxic T cells peptides that are derived mostly from cytosolic

proteins, which are ubiquitinated and then degraded by the proteasome^{3,4}. Here we show that on inflammatory stimulation, DCs accumulate newly synthesized ubiquitinated proteins in large cytosolic structures. These structures are similar to, but distinct from, aggresomes and inclusion bodies observed in many amyloid diseases^{5,6}. Notably, these dendritic cell aggresome-like induced structures (DALIS) are transient, require continuous protein synthesis and do not affect the ubiquitin–proteasome pathway. Our observations suggest the existence of an organized prioritization of protein degradation in stimulated DCs, which is probably important for regulating MHC class I presentation during maturation.

MHC class I presentation is linked directly to proteasome activity and consequently to cellular ubiquitinated protein levels⁷. To determine whether a principal regulation of MHC class I antigen presentation exists during DC maturation, we investigated whether protein ubiquitination was affected in these cells. We determined total protein ubiquitination levels in Langerhans cells (LCs) by confocal microscopy using the FK2 monoclonal antibody. FK2 is specific for mono- or polyubiquitinated proteins and, importantly,

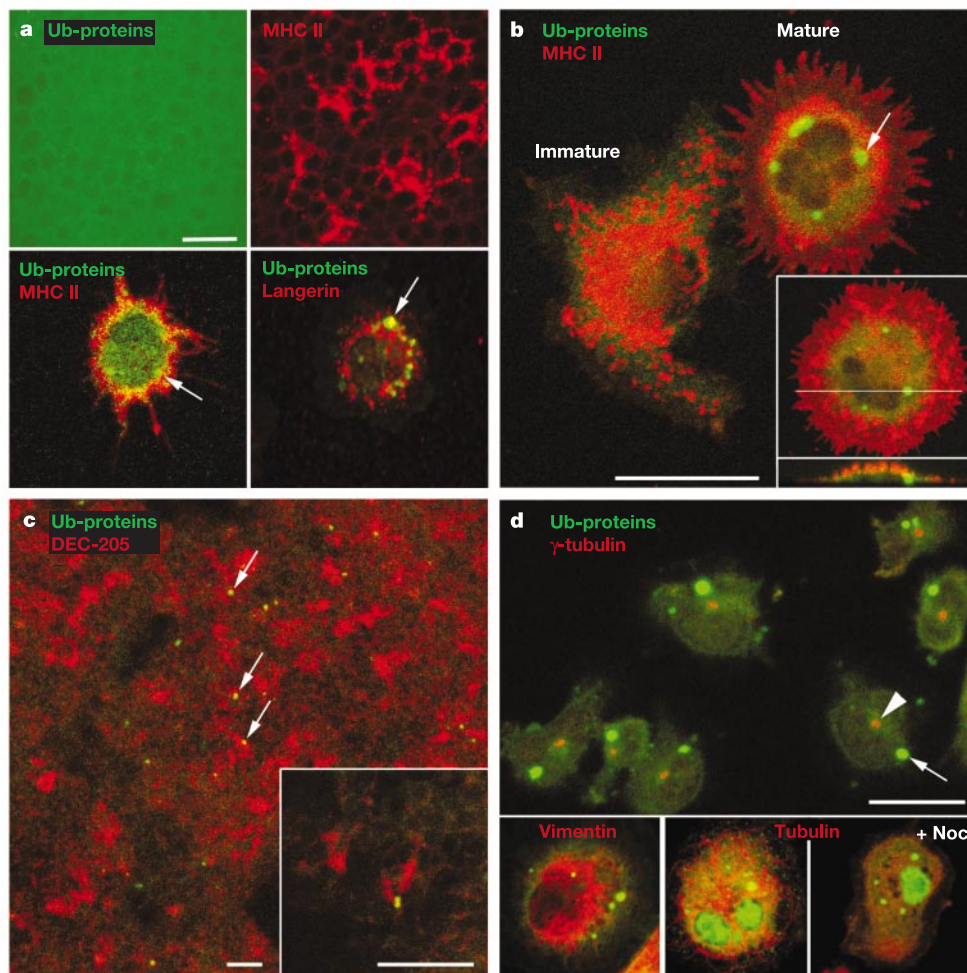


Figure 1 Ubiquitinated protein accumulation during DC maturation. **a**, Explanted mouse epidermal sheets (top row) and emigrated Langerhans cells (bottom row) stained for ubiquitin conjugates (green), MHC class II or langerin (red), and visualized by confocal microscopy. Upon LPS treatment, LCs accumulated ubiquitinated proteins (Ub-proteins) in peripheral aggregates (arrows). **b**, Maturing human CD34⁺-derived LCs aggregated ubiquitinated proteins (green, indicated by an arrow), but not immature LCs, expressing HLA-DR (red) in lysosomal compartments. Z-sectioning revealed the circular shape of DALISs. **c**, In sections of mouse lymph node, only DCs (DEC-205 staining, in red)

accumulated ubiquitinated protein aggregates (green, indicated by arrows). Inset is a higher magnification image. **d**, Ubiquitinated protein aggregates (green; arrow) did not co-localize with the MTOC (γ -tubulin, in red and indicated by an arrowhead) in maturing mouse DCs derived from bone marrow, nor induce vimentin caging (red) in human LCs. Attempts to destabilize DALISs with nocodazole (Noc) were unsuccessful. The methanol fixation used to preserve microtubules (red) revealed nuclear ubiquitinated protein (large green patches). Scale bars, 20 μ m.

does not react with free ubiquitin⁸. We detected only weak cytosolic FK2 staining in mouse epidermal sheets (Fig. 1a), with no obvious co-localization with resident immature LCs. In contrast, we observed a strong accumulation of ubiquitinated proteins in mature emigrated LCs (Fig. 1a). Of note, in addition to slightly increased levels of soluble ubiquitinated substrates, FK2 staining revealed a strong peripheral aggregation of ubiquitinated material in most of the mature LCs, as shown by high cell-surface expression of MHC class II. We confirmed this observation in mature human CD34⁺ precursor-derived LCs⁹ in which ubiquitinated aggregates have a smooth and regular spherical shape and can reach a diameter of greater than 4 μm (Fig. 1b and Z section). Most LCs contained only one principal aggregate, but their number could reach five per cell, a result also obtained with DCs derived from mouse bone marrow¹⁰.

In cryostat sections of mouse lymph node, most of the aggregate-containing cells also expressed the DC-specific lectin DEC-205 (ref. 11) (Fig. 1c). Aggregation of ubiquitinated proteins is therefore probably restricted to DCs and occurs during maturation both *in*

vitro and *in vivo*. Staining of other cell types (for example, B cells) did not show any aggregate formation on treatment with lipopolysaccharide (LPS), suggesting further the DC specificity of the phenomenon (not shown).

Several examples of pathological aggregation of ubiquitinated proteins characteristic of amyloid diseases have been reported⁵. This phenomenon is also observed *in vitro* on pharmacological inhibition of the proteasome^{12–14} or in cells transfected with mutated complementary DNAs¹⁵. These aggregates have been called aggresomes⁵ and result from microtubule-dependent conglomeration of smaller aggregates in the area of the microtubule-organizing centre (MTOC)^{14,15}. Aggresomes are surrounded by a cage formed by the intermediate filament protein vimentin^{14,15}. As shown in Fig. 1d, FK2-positive aggregates did not co-localize with γ -tubulin (MTOC) nor vimentin. Dendritic-cell aggregate formation was not affected by the microtubule-depolymerizing drug nocodazole, indicating that these structures are distinct from 'classical' aggresomes observed under pathological circumstances. To underline this dis-

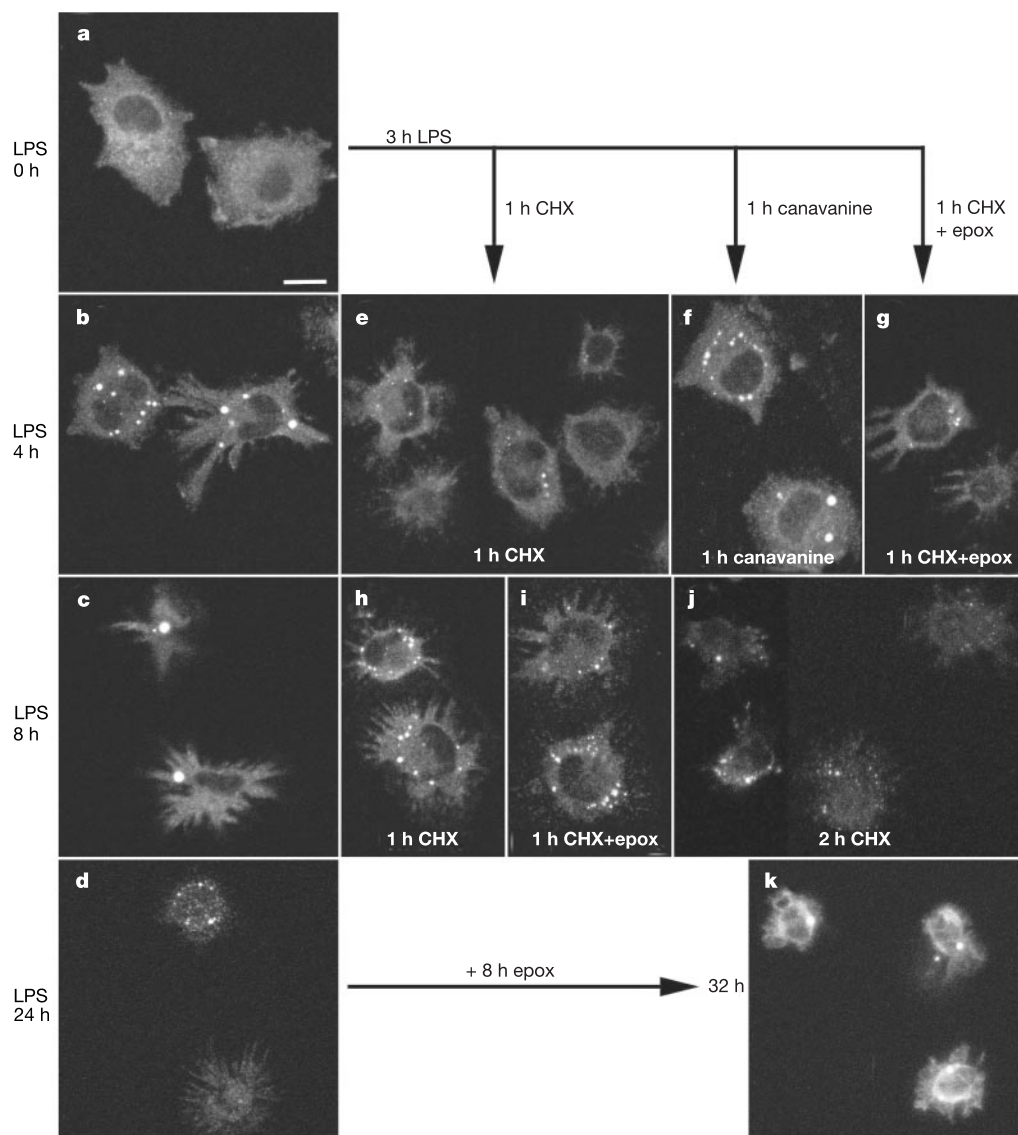


Figure 2 DALIS formation is transient and depends on protein translation. Confocal microscopy was performed on maturing DCs derived from bone marrow to follow the kinetics of DALIS formation. **a–d**, DALIS appeared as early as 4 h after LPS stimulation, reached peak size at 8 h, and diminished in numbers and size at 24 h. **e–j**, One hour of cycloheximide (CHX) treatment prevented DALIS formation (**e**) or induced their

fragmentation at their peak size (**h, j**), whereas treatment with canavanine enhanced DALIS formation (**f**). Proteasome inhibition with epoxomicin (epox) did not counteract CHX treatment (**g, i**). **k**, At 24 h of maturation, incubation for a further 8 h with epoxomicin prevented DALIS disappearance. Scale bar, 10 μm .

tion, we name these aggregates 'DALIS', for dendritic cell aggresome-like inducible structures.

DALIS appeared as early as 4 h and peaked at around 8 h after LPS stimulation (Fig. 2a–c). After 24–36 h, most of the mature DCs no longer contained DALIS, and the average size of the remaining structures strongly diminished (Fig. 2d). DALIS are therefore transient, contrary to what has been observed for aggresomes.

An extremely high proportion of newly synthesized proteins represent defective ribosomal products (DRiPs), which are ubiquitinated and degraded shortly after synthesis¹⁶. DRiPs could therefore be an important part of material forming DALIS. We evaluated the contribution of newly synthesized proteins to DALIS formation using cycloheximide (CHX). A 1-h CHX treatment on maturing DCs prevented DALIS formation (Fig. 2e). Interestingly, inhibition of protein synthesis at a later maturation time (8 h) induced rapid DALIS fragmentation and often total disappearance after 2 h (Fig. 2h, j). The same happened also on anisomycin or emetine treatment, both of which are potent inhibitors of translation¹⁷ (not shown). Protein synthesis is therefore required for the formation and the maintenance of the aggregate. To test whether DRiPs are

incorporated in DALIS, we incubated maturing DCs with the arginine analogue, canavanine. Canavanine causes synthesis of abnormal proteins (DRiPs), but unlike CHX, it does not inhibit translation¹³. Canavanine stimulated DALIS formation, indicating that aggregation requires functional translation but not necessarily intact proteins (Fig. 2f).

We used epoxomicin, a potent proteasome inhibitor¹⁸, to determine the contribution of proteolysis to DALIS fragmentation. Epoxomicin did not interfere with the disappearance of DALIS induced by CHX at early stages of maturation (Fig. 2g, i). The proteasome, which is inhibited after such treatment (Fig. 3), is therefore not involved directly in pharmacological destabilization of DALIS. On the contrary, epoxomicin treatment of mature DCs (24 h LPS) strongly enhanced accumulation of ubiquitinated proteins and re-formation of DALIS (Fig. 2k). The proteasome, therefore, participates actively in DALIS disappearance during late stages of maturation.

It has been suggested that aggresomes impair the ubiquitin–proteasome pathway, with potential cytotoxic effects in the case of neurodegenerative diseases¹⁹. To visualize a potential interference of

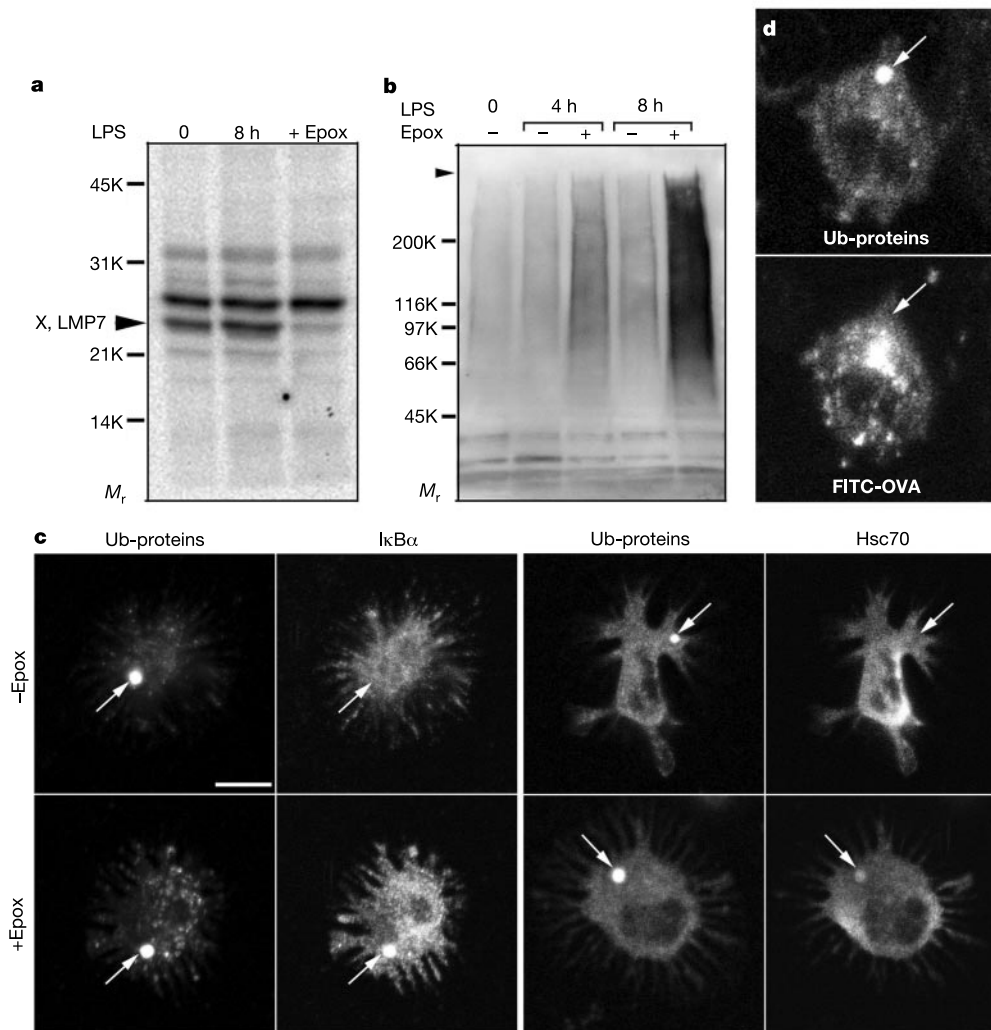


Figure 3 Proteasome activity is not affected during DALIS formation and DCs can discriminate among ubiquitinated proteins. **a**, Proteasome activity, assessed by the active-site-directed probe ¹²⁵I-labelled NLVS, was similar in immature (lane 1) and maturing DCs (8 h LPS; lane 2). Pre-treatment for 1 h with epoxomicin inhibited subunits X and LMP7 of the proteasome (lane 3). **b**, Total ubiquitinated proteins, visualized by western blot with FK2, increased slightly during DC maturation. Proteasome inhibition

with epoxomicin led to the massive accumulation of ubiquitinated proteins. **c**, IκBα was not enriched in DALIS (arrows), except when epoxomicin was added (bottom of second column). The same was observed for Hsc70, a chaperone normally found in aggresomes (far right column). Scale bar, 10 μm. **d**, Endocytosed FITC-ovalbumin was not enriched in DALIS after cytoplasmic translocation induced by CD40 ligation. Scale 1.4 × that in **c**.

DALISs with proteolysis, proteasome activity was evaluated by active site radioactive labelling. Incubation of DCs with the proteasome-specific probe ^{125}I -labelled NLVS²⁰ (Fig. 3a) did not reveal any decrease in proteasome activity during maturation, contrary to the decreased activity observed on epoxomicin treatment. Furthermore, ubiquitinated protein accumulation in DCs treated with LPS for 4–8 h was strongly enhanced by epoxomicin, although LPS treatment alone provoked only a limited increase in ubiquitinated protein levels, as seen by immunoblotting (Fig. 3b). In conclusion, the proteasome seems to be functional in maturing DCs, and DALIS formation does not depend on nor seem to affect deeply its activity.

Efficient nuclear translocation of the transcription factor NF- κB , a necessary component of DC maturation²¹, requires ubiquitination and degradation of the phosphorylated I $\kappa\text{B}\alpha$ chaperone protein. I $\kappa\text{B}\alpha$ is therefore well suited to study the intracellular fate of ubiquitinated proteins during DC maturation. After 8 h of LPS stimulation, we did not observe any enrichment of I $\kappa\text{B}\alpha$ in DALIS (Fig. 3c). Notably, pharmacological inhibition of the proteasome promoted accumulation in DALIS of I $\kappa\text{B}\alpha$ as well as the heat shock protein Hsc70 (Fig. 3c). I $\kappa\text{B}\alpha$ is therefore normally degraded in maturing DCs, which seem to be able to discriminate among different ubiquitinated proteins during DALIS formation. Aggregated proteins are likely to be more resistant to proteolysis, and DCs could prioritize selected ubiquitinated protein degradation (for

example, I κB) by not storing them in DALIS. However, DALIS aggregation specificity can be abrogated by epoxomicin treatment. The exclusion from DALIS of Hsc70, a molecule normally found in aggresomes⁵, further supports the distinction between these two types of structure.

We investigated the fate of endocytosed proteins during translocation from the endocytic pathway to the cytosol. This transport pathway seems to be DC specific and is required for cross-presentation²². After fluorescein isothiocyanate (FITC)-conjugated ovalbumin uptake, DCs were activated with CD40 ligation to induce maturation. Although under these conditions DALIS formation occurred normally, no selective enrichment of cytosolic FITC-ovalbumin was detected (Fig. 3d). Translocated proteins are therefore not specifically targeted to the aggregates, further supporting the existence of selective protein degradation during early DC maturation.

FK2 detects a large pool of cytosolic and nuclear ubiquitinated proteins (Figs 1d and 3b). We adapted a nuclear matrix protein isolation procedure²³ to partially purify DALIS by extracting contaminating cytosolic and nuclear ubiquitinated material. DALIS resisted sequential extractions with Triton-X100, DNase I and 2 M NaCl, whereas most other ubiquitin-conjugated material was eliminated under such conditions (see Supplementary Information). We used the FK1 antibody, specific for polyubiquitinated proteins⁸, to demonstrate that DALIS contain polyubiquitinated conjugates,

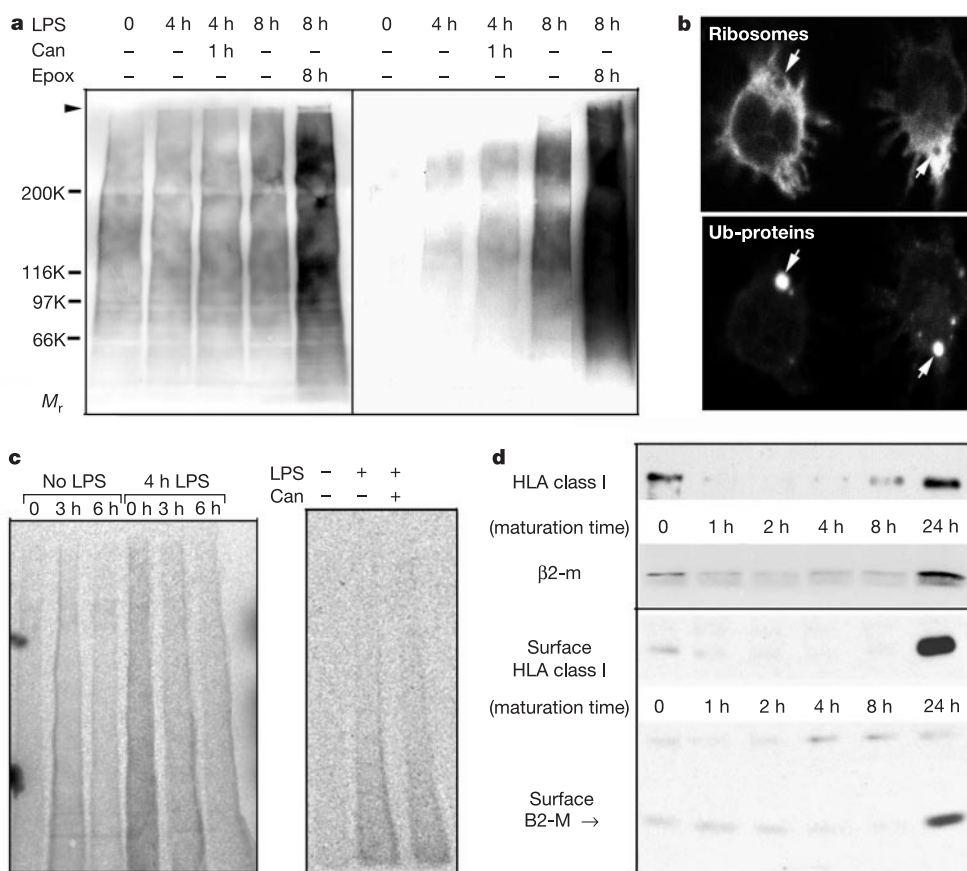


Figure 4 Newly synthesized proteins are targeted to DALIS. **a**, Immature, maturing (4 h and 8 h LPS), canavanine- (can), or epoxomicin-treated DCs were submitted to biochemical extraction, SDS-PAGE and FK2 immunolabelling. Triton-X100 soluble fractions are shown on the left and 2 M NaCl non-extractable material enriched for DALIS are on the right. **b**, Ribosomes (top) are enriched in the vicinity of DALIS (bottom) in maturing DCs. Scale is the same as in Fig. 3c. **c**, Immature and maturing DCs with or without stimulation by LPS for 4 h were pulse-labelled with ^{35}S -labelled Promix (left) or

^{35}S -labelled cysteine (right) for 5 min, and chased for the indicated times, before selective extraction and FK2 immunoprecipitation. **d**, Immunoprecipitation from human LCs of HLA class I with W6/32 (top), or surface biotinylated complexes with B9.12.1 (bottom) at different times of maturation, induced either by TNF- α (top) or bacteria (bottom). Immunoprecipitated material (heavy chain and $\beta 2$ -microglobulin; $\beta 2\text{-m}$) was revealed by immunoblot with specific antibodies or anti-biotin.

whereas most nuclear ubiquitinated proteins are monoconjugated¹³. Interestingly, the level of detergent-soluble ubiquitinated proteins detected by immunoblot was unaffected by DC maturation, suggesting that a large increase in ubiquitinated material does not promote DALIS formation (Fig. 4a, left). After selective enrichment, ubiquitinated proteins were found only in the insoluble material of maturing DCs (Fig. 4a, right). The smeary appearance and the overall molecular mass pattern indicated that an array of proteins is incorporated in DALIS. Epoxomicin treatment induced abnormal accumulation of ubiquitinated proteins both in soluble and insoluble DC fractions. Importantly, induction of DRiPs by a 1-h canavanin treatment significantly increased the abundance of insoluble ubiquitinated material in maturing DCs, whereas the Triton-X100 soluble pool was unchanged (Fig. 4a). This last piece of evidence confirms that DRiPs are incorporated into DALIS.

To further investigate DRiP incorporation into DALIS, we performed radioactive pulse-chase experiments. After selective extraction, the ubiquitinated material was immunoprecipitated with FK2 before separation and autoradiography (Fig. 4c, left). High levels of insoluble radioactivity were found already in maturing DCs after a short pulse (5 min), indicating that newly synthesized proteins were immediately ubiquitinated and rapidly targeted to DALIS. Furthermore, immunoprecipitated ubiquitinated proteins were relatively long lived as they could still be detected after 6 h of chase. As expected, immature DCs did not display this type of rapid aggregation; however, radioactive ubiquitinated proteins could be immunoprecipitated from insoluble fractions after 3 h of chase. At that time, maturation and subsequent DALIS formation were induced by ³⁵S-labelled Promix, most probably contaminated by endotoxin (not shown). To rule out the possibility that the observed labelling could be due to incorporation of ³⁵S into ubiquitin only, a 5-min pulse was performed with ³⁵S-labelled cysteine, an amino acid absent from the ubiquitin molecule (Fig. 4c, right). Ubiquitinated proteins tagged with ³⁵S-labelled cysteine aggregated rapidly in maturing DCs, demonstrating that newly translated proteins, conjugated to pre-existing ubiquitin, can form DALIS. In agreement with the DRiP hypothesis, aggregation of radioactive proteins also occurred in maturing DCs pre-treated with canavanin.

The extreme rapidity with which newly synthesized proteins are incorporated into DALIS led us to investigate whether the ribosomal machinery was affected during DALIS formation. Ribosomes, detected by an antibody specific for the S6 ribosomal protein, were enriched in rings surrounding the aggregates. In extreme cases, the translation apparatus was almost completely redistributed in the vicinity of DALIS (Fig. 4b). This marked phenomenon could explain the rapidity of DRiP incorporation into DALIS, and suggests that it might happen along with translation.

The direct link between protein translation, DRiPs and MHC class-I-restricted presentation has been demonstrated by several laboratories using different techniques^{16,24,25}. DRiPs are considered as the main source for MHC class-I-restricted antigenic peptides in both professional (that is, DCs, B cells, macrophages) and non-professional (that is, epithelial/endothelial cells) antigen-presenting cells. We therefore investigated MHC class I-peptide complex assembly in maturing LCs, on which DALIS occurrence should have an impact. W6/32 and B9.12.1 are antibodies used to detect assembled human heterotrimeric complexes formed by β 2-microglobulin, class I heavy chain and peptide²⁶. The amount of immunoprecipitated molecules reflects the activity of the HLA class I presentation pathway at a given time. In maturing CD34⁺-derived human LCs, class I heavy chain and β 2-microglobulin immunoprecipitated by W6/32 were lost rapidly, before a major increase after 24 h (Fig. 4d, top). This late increase was probably due to enhanced transcription and translation of *HLA* genes²⁷. The rapid loss of reactivity indicated a high turnover of the class I complexes and an overall reduction of its formation on stimulation. Loss of

W6/32 epitopes correlated strongly with the kinetics of DALIS formation (Fig. 2). HLA class I assembly and loading seem therefore dependent on DRiP storage, and probably result from a biochemical event occurring very rapidly after stimulation. This regulation also involves the rapid disassembly of pre-existing surface heteromeric complexes, as shown by immunoprecipitation of biotinylated surface molecules with B9.12.1 in bacteria-stimulated LCs (Fig. 4d, bottom).

Until now, observation of large protein aggregates *in vivo* was limited to pathological situations. We show here that reversible ubiquitinated protein aggregation occurs in a normal physiological context. We demonstrate that this aggregation is DC specific and inducible by inflammatory stimuli. DALIS formation could be the result of a massive increase in the total level of ubiquitination during maturation with a saturating effect on the proteasome. Our data do not support this hypothesis, however, as little difference in the total ubiquitinated protein pool is found between immature and maturing DCs. Furthermore, if DALIS were the result of a mere saturation of the proteasome-ubiquitin degradation pathway, molecules such as I κ B α should be enriched in these structures as observed when proteolysis is inhibited with epoxomicin. Although the mechanism of DALIS formation is still unclear, we have determined that a constant flow of newly synthesized proteins, likely to be DRiPs, supports the growth of these structures. The relatively long half-life of insoluble polyubiquitinated proteins and the constant enlargement of DALIS for at least 8 h suggest that DRiPs are poorly degraded at the onset of DC maturation. The speed of the ubiquitination and aggregation process implies the existence of a specific sorting machinery dedicated to DALIS formation, to which ribosomes might physically participate.

By limiting DRiP degradation, maturing DCs might have the ability to control MHC class I loading and peptide presentation²⁸. This hypothesis is strongly supported by the rapid decrease in W6/32 and B9.12.1 epitopes observed at early stages of maturation. This drop probably reflects a change in peptide availability for MHC loading, with potential consequences on antigen presentation by DCs, as well as on their role in establishing tolerance and immunity²⁹. DALIS could also represent antigen storage structures. The precise timing of DALIS formation could allow the coordination of MHC class I and class II processing at DC arrival in lymph nodes. □

Methods

Materials and cell culture

We purchased all chemicals from Sigma unless stated otherwise. Epoxomicin was purchased from Affinity.

Male C57/BL6 mice (7–8 weeks of age) were purchased from Charles River Laboratories. Bone-marrow-derived DCs and epidermal sheets were cultured as described previously¹⁰. Human Langerhans cells were obtained as described⁹ and matured with tumour-necrosis factor- α (Preprotech) or *Escherichia coli*.

Antibodies and immunocytochemistry

We performed immunofluorescence confocal microscopy with Leica TCS 4D as described¹⁰. Murine I-A was detected using the 'Rivoli' affinity-purified rabbit polyclonal antibody¹⁰. Human HLA class I and II were detected using anti-heavy-chain and anti-DR rabbit polyclonal antibody, respectively (gift of J. Neefjes). Polyclonal anti-tubulin and γ -tubulin were a gift of T. Kreis and M. Bornens. Rat monoclonal NLDC-145 anti-DEC205 was obtained from R. Steinman, and rabbit polyclonal anti-langerin was a gift of S. Sealand. Monoclonal antibodies FK1, FK2 and polyclonal antibody against ubiquitin were obtained from Affinity. Anti-I κ B α and anti-Hsc70 were obtained from Santa-Cruz. Anti-S6 ribosomal protein was obtained from Cell Signaling Technology. Anti- β 2-microglobulin, W6/32 and B9.12.1 were from Biotest. All secondary antibodies were from Immunotech. For pharmacological treatments, cells were incubated respectively with 1 μ M epoxomicin, 15 mM canavanin or 25 μ M CHX. Dendritic cells were incubated with FITC-OVA (10 μ g ml⁻¹) for 4 h before CD40 ligation with rat anti-mouse CD40 antibody (FSK45) for an additional 4 h.

Proteasome catalytic subunit labelling

Specific proteasome active-site inhibitor NLVS (gift of H. Ploegh) was iodinated using the iodogen method as reported previously²⁰. Immature and maturing DCs (7 h LPS), pre-treated or not with 1 μ M epoxomicin (1 h), were incubated for 1 h with ¹²⁵I-labelled NLVS at 37 °C, and washed twice with PBS before lysis in sample buffer and 15% SDS-polyacrylamide gel electrophoresis analysis.

Selective enrichment of DALIS

Nocodazole-treated DCs were submitted to a 30-min 1% Triton-X100 extraction at 4 °C before a 30-min DNase I (100 µg ml⁻¹) treatment at 37 °C in 20 mM Tris pH 7.4, 1.5 mM MgCl₂, 10 mM NaCl. Nuclear material extraction was performed for 30 min at 4 °C with a solution of 20 mM Tris pH 7.4, 1.5 mM MgCl₂, 2 M NaCl. Remaining material was washed with PBS and submitted to immunocytochemical and biochemical analysis.

Radiolabelling, immunoprecipitation and immunoblots

Roughly 2×10^7 DCs were pulse-labelled with 10 mCi ml⁻¹ ³⁵S-labelled Promix or ³⁵S-labelled cysteine (both from Amersham Pharmacia Biotech) in labelling medium for 5 min, and chased for various time at 37 °C as described¹⁰. Radiolabelled DALIS-enriched samples were solubilized at 95 °C for 5 min in SDS 1% before dilution in 10 mM Tris, 150 mM NaCl (pH 7.4) containing 1% Triton-X100, 1 µM epoxomicin and protease inhibitor cocktail. Samples were pre-cleared with protein A Sepharose (Amersham Pharmacia Biotech) at 4 °C and immunoprecipitated overnight with the FK2 antibody. Immunoprecipitates were analysed onto 2–10% gradient SDS–PAGE gels and quantified with a Fuji phosphorimager. Immunoprecipitations with W6/32 and B9.12.1 (after surface biotinylation) were performed as described¹⁰ for 2 h at 4 °C before immunoblot analysis. Immunoblots were performed after transfer on Immobilon P (Millipore) and detected by chemiluminescence (Amersham Pharmacia Biotech).

Received 28 December 2001; accepted 19 February 2002.

- Banchereau, J. & Steinman, R. M. Dendritic cells and the control of immunity. *Nature* **392**, 245–252 (1998).
- Mellman, I. & Steinman, R. M. Dendritic cells: specialized and regulated antigen processing machines. *Cell* **106**, 255–258 (2001).
- Hirsch, C. & Ploegh, H. L. Intracellular targeting of the proteasome. *Trends Cell Biol.* **10**, 268–272 (2000).
- Yewdell, J. W. & Bennink, J. R. Cut and trim: generating MHC class I peptide ligands. *Curr. Opin. Immunol.* **13**, 13–18 (2001).
- Kopito, R. R. Aggregates, inclusion bodies and protein aggregation. *Trends Cell Biol.* **10**, 524–530 (2000).
- Sherman, M. Y. & Goldberg, A. L. Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases. *Neuron* **29**, 15–32 (2001).
- Weissman, A. M. Themes and variations on ubiquitylation. *Nature Rev. Mol. Cell. Biol.* **2**, 169–178 (2001).
- Fujimuro, M., Sawada, H. & Yokosawa, H. Production and characterization of monoclonal antibodies specific to multi-ubiquitin chains of polyubiquitinated proteins. *FEBS Lett.* **349**, 173–180 (1994).
- Gatti, E. *et al.* Large-scale culture and selective maturation of human Langerhans cells from granulocyte colony-stimulating factor-mobilized CD34⁺ progenitors. *J. Immunol.* **164**, 3600–3607 (2000).
- Pierre, P. *et al.* Developmental regulation of MHC class II transport in mouse dendritic cells. *Nature* **388**, 787–792 (1997).
- Jiang, W. *et al.* The receptor DEC-205 expressed by dendritic cells and thymic epithelial cells is involved in antigen processing. *Nature* **375**, 151–155 (1995).
- Wojcik, C., Schroeter, D., Wilk, S., Lamprecht, J. & Pawletz, N. Ubiquitin-mediated proteolysis centers in HeLa cells: indication from studies of an inhibitor of the chymotrypsin-like activity of the proteasome. *Eur. J. Cell Biol.* **71**, 311–318 (1996).
- Anton, L. C. *et al.* Intracellular localization of proteasomal degradation of a viral antigen. *J. Cell. Biol.* **146**, 113–124 (1999).
- Johnston, J. A., Ward, C. L. & Kopito, R. R. Aggregates: a cellular response to misfolded proteins. *J. Cell. Biol.* **143**, 1883–1898 (1998).
- Garcia-Mata, R., Bebek, Z., Sorscher, E. J. & Szul, E. S. Characterization and dynamics of aggregate formation by a cytosolic GFP-chimera. *J. Cell. Biol.* **146**, 1239–1254 (1999).
- Schubert, U. *et al.* Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* **404**, 770–774 (2000).
- Iordanov, M. S. *et al.* Ribotoxic stress response: activation of the stress-activated protein kinase JNK1 by inhibitors of the peptidyl transferase reaction and by sequence-specific RNA damage to the alpha-sarcosine loop in the 28S rRNA. *Mol. Cell. Biol.* **17**, 3373–3381 (1997).
- Meng, L. *et al.* Epoxomicin, a potent and selective proteasome inhibitor, exhibits *in vivo* anti-inflammatory activity. *Proc. Natl Acad. Sci. USA* **96**, 10403–10408 (1999).
- Bence, N. F., Sampat, R. M. & Kopito, R. R. Impairment of the ubiquitin-proteasome system by protein aggregation. *Science* **292**, 1552–1555 (2001).
- Bogyo, M., Shin, S., McMaster, J. S. & Ploegh, H. L. Substrate binding and sequence preference of the proteasome revealed by active-site-directed affinity probes. *Chem. Biol.* **5**, 307–320 (1998).
- Rescigno, M., Martino, M., Sutherland, C. L., Gold, M. R. & Ricciardi-Castagnoli, P. Dendritic cell survival and maturation are regulated by different signalling pathways. *J. Exp. Med.* **188**, 2175–2180 (1998).
- Thery, C. & Amigorena, S. The cell biology of antigen presentation in dendritic cells. *Curr. Opin. Immunol.* **13**, 45–51 (2001).
- Staufenbiel, M. & Deppert, W. Preparation of nuclear matrices from cultured cells: subfractionation of nuclei *in situ*. *J. Cell Biol.* **98**, 1886–1894 (1984).
- Reits, E. A., Vos, J. C., Gromme, M. & Neefjes, J. The major substrates for TAP I are derived from newly synthesized proteins. *Nature* **404**, 774–778 (2000).
- Khan, S. *et al.* Neosynthesis is required for the presentation of a T cell epitope from a long-lived viral protein. *J. Immunol.* **167**, 4801–4804 (2001).
- Brodsky, F. M. & Parham, P. Monoclonal anti-HLA-A,B,C monoclonal antibodies detecting molecular subunits and combinatorial determinants. *J. Immunol.* **128**, 129–135 (1982).
- Huang, Q. *et al.* The plasticity of dendritic cell responses to pathogens and their components. *Science* **294**, 870–875 (2001).
- Yewdell, J. W., Schubert, U. & Bennink, J. R. At the crossroads of cell biology and immunology: DRiPs and other sources of peptide ligands for MHC class I molecules. *J. Cell. Sci.* **114**, 845–851 (2001).
- Hawiger, D. *et al.* Dendritic cells induce peripheral T cell unresponsiveness under steady state conditions *in vivo*. *J. Exp. Med.* **194**, 769–779 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank A.-M. Imbert, C. Chabannon and all the personnel at the cell repository of the Centre de Thérapie Cellulaire (Institut Paoli-Calmettes) for providing access to human CD34⁺ cells. We also thank J. Yewdell, P. Machy and L. Delamarre for a gift of antibody and for discussions. This work is supported by grants to P.P. from the Ministère de la Recherche et de la Technologie (Action Concertée Initiative Blanche), the Association pour la Recherche contre le Cancer, and the Fondation Schlumberger pour l'Éducation et la Recherche. H.L. is supported by a Sidaction and Agence Nationale de Recherches sur la Sida fellowships, and E.G. is an EU Marie Curie fellow.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.P. (e-mail: pierre@ciml.univ-mrs.fr).

Identification of a factor that links apoptotic cells to phagocytes

Rikinari Hanayama^{*†}, Masato Tanaka^{*†}, Keiko Miwa^{*†}, Azusa Shinohara[‡], Akihiro Iwamatsu[‡] & Shigekazu Nagata^{*†}

^{*} Department of Genetics, Osaka University Medical School, [†] Core Research for Evolutional Science and Technology, Japan Science and Technology Corporation, 2-2 Yamada-oka, Suita, Osaka, 565-0871, Japan

[‡] Central Laboratories for Key Technology, Kirin Brewery Co., Ltd, 1-13-5 Fukuura, Kanazawa, Yokohama, Kanagawa 236-0004, Japan

Apoptotic cells are rapidly engulfed by phagocytes to prevent the release of potentially noxious or immunogenic intracellular materials from the dying cells, thereby preserving the integrity and function of the surrounding tissue¹. Phagocytes engulf apoptotic but not healthy cells, indicating that the apoptotic cells present a signal to the phagocytes, and the phagocytes recognize the signal using a specific receptor². Here, we report a factor that links apoptotic cells to phagocytes. We found that milk fat globule-EGF-factor 8 (MFG-E8)^{3,4}, a secreted glycoprotein, was produced by thioglycollate-elicited macrophages. MFG-E8 specifically bound to apoptotic cells by recognizing aminophospholipids such as phosphatidylserine. MFG-E8, when engaged by phospholipids, bound to cells via its RGD (arginine-glycine-aspartate) motif—it bound particularly strongly to cells expressing $\alpha_v\beta_3$ integrin. The NIH3T3 cell transformants that expressed a high level of $\alpha_v\beta_3$ integrin were found to engulf apoptotic cells when MFG-E8 was added. MFG-E8 carrying a point mutation in the RGD motif behaved as a dominant-negative form, and inhibited the phagocytosis of apoptotic cells by peritoneal macrophages *in vitro* and *in vivo*. These results indicate that MFG-E8 secreted from activated macrophages binds to apoptotic cells, and brings them to phagocytes for engulfment.

Cells expressing a caspase-resistant ICAD (inhibitor of caspase-activated DNase) do not undergo apoptotic DNA fragmentation, but their DNA can still be cleaved when the cells are phagocytosed by macrophages⁵. This system was used to examine the phagocytosis of apoptotic cells. As shown in Fig. 1a, when thymocytes from transgenic mice carrying caspase-resistant ICAD mutant (ICAD-Sdm)⁵ were treated with dexamethasone, about 50% of the cells became annexin-V-positive within 4 h, but they were not stained by TUNEL (TdT-mediated dUTP nick end labelling). Co-culture of macrophages with apoptotic thymocytes, but not with freshly

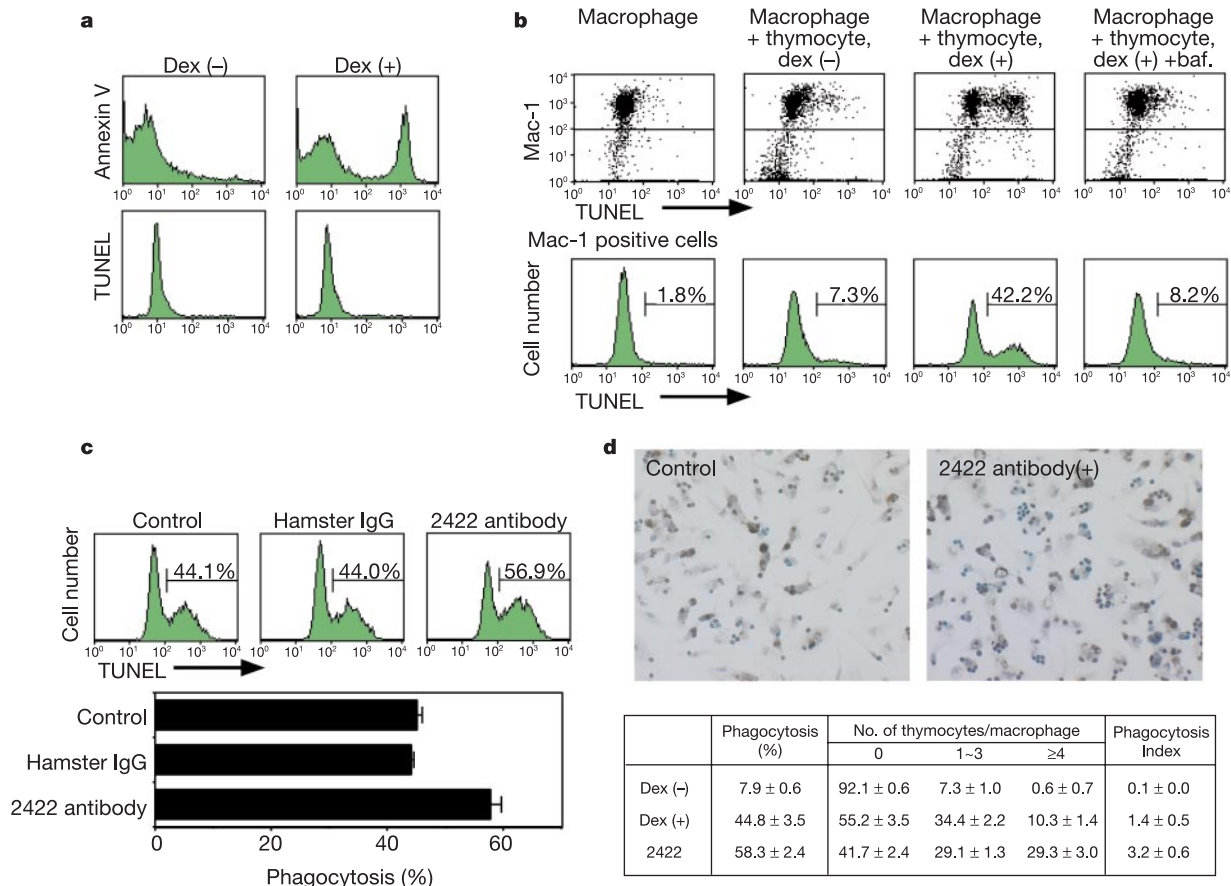


Figure 1 Establishment of monoclonal antibody that enhances the phagocytosis of apoptotic cells. **a**, Thymocytes from ICAD-Sdm mice were untreated or treated with dexamethasone for 4 h, and stained with PE-conjugated annexin V (BD Pharmingen) or TUNEL. **b**, Thioglycollate-elicited mouse peritoneal macrophages were incubated with thymocytes from ICAD-Sdm mice that had been treated (+) or not treated (-) with dexamethasone. The cells were then stained for Mac-1 and TUNEL. Bafilomycin (100 nM) was added to macrophages 30 min before incubation with thymocytes. Lower panels show the TUNEL-staining profiles in the Mac-1⁺ population with the percentage of TUNEL-positive cells. **c**, Phagocytosis was assayed in the absence or presence of

12 $\mu\text{g ml}^{-1}$ normal hamster IgG, or the 2422 antibody, and the FACS profiles of TUNEL-staining in Mac-1⁺ population are shown. Assay in triplicate was carried out three times, and the average values with s.d. are shown in the lower panel. **d**, Phagocytosis was assayed in the absence or presence of the 2422 antibody, stained for TUNEL, and observed by light microscopy (upper panel). Magnification, $\times 200$. The experiments were done 3 times. The percentage of macrophages that carried none, 1-3, or ≥ 4 TUNEL-positive cells was determined, and shown in lower panel. Per cent phagocytosis indicates the percentage of macrophages carrying apoptotic thymocytes. Phagocytosis index indicates the number of apoptotic cells per macrophages.

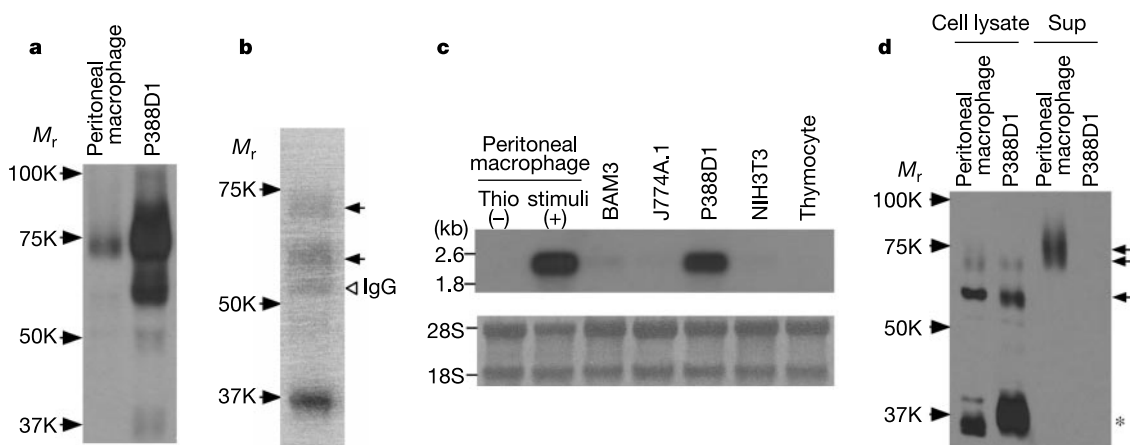


Figure 2 Identification of MFG-E8, and its expression. **a**, Thioglycollate-elicited peritoneal macrophages or P388D1 cells were surface-labelled with biotin, immunoprecipitated with the 2422 antibody, and analysed by western blotting with peroxidase-conjugated streptavidin. **b**, Proteins recognized by the 2422 antibody were purified from P388D1 cells, separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE), and stained with ponceau-S. Arrows at right indicate proteins subjected to protein sequence analysis. **c**, RNA (7.5 μg) from the indicated cells was subjected to northern hybridization using

³²P-labelled murine MFG-E8 cDNA (upper panel). In lower panel, the filter was stained with methylene blue. **d**, Thioglycollate-elicited peritoneal macrophages and P388D1 were cultured for 48 h. The cell lysates and culture supernatants were immunoprecipitated with 2422 antibody, and western blotted with rabbit anti-MFG-E8 antibodies. Arrows indicate MFG-E8 proteins. Bands indicated by an asterisk are probably degraded products of MFG-E8.

prepared thymocytes, rendered about 40% of the Mac-1⁺ cells TUNEL-positive (Fig. 1b). Treating macrophages with bafilomycin, which prevents the acidification of lysosomes⁶, inhibited the appearance of TUNEL-positive macrophages. These results indicated that the macrophages specifically engulfed apoptotic cells and digested their chromosomal DNA. To identify mediators of this process, Armenian hamsters were immunized with thioglycollate-elicited mouse peritoneal macrophages, and hybridomas were prepared. Among several thousand hybridomas, one monoclonal antibody (2422) increased the percentage of macrophages that engulfed the apoptotic cells from 44% to 57% (Fig. 1c). Microscopic observations also indicated that the number of engulfed apoptotic cells per macrophage increased from 1.4 to 3.2 in the presence of the 2422 antibody (Fig. 1d).

To identify the protein recognized by the 2422 antibody, thioglycollate-elicited peritoneal macrophages and a macrophage cell line P388D1 were surface-labelled with biotin, and subjected to immunoprecipitation with this antibody. Western blotting of the immunoprecipitates with peroxidase-conjugated streptavidin revealed bands of relative molecular mass 72,000 and 56,000 (M_r 72K and 56K; Fig. 2a). The proteins in P388D1 were affinity-purified with the 2422 antibody (Fig. 2b), and a mass spectrometry analysis of peptides from these proteins identified them as mouse MFG-E8 (refs. 3, 4). Two classes of complementary DNA (MFG-E8-L and MFG-E8-S) were isolated from mouse peritoneal macrophages by polymerase chain reaction. MFG-E8 messenger RNA was expressed in thioglycollate-elicited peritoneal macrophages and P388D1 (Fig. 2c). In contrast, little MFG-E8 mRNA was detected in resident peritoneal macrophages, macrophage cell lines such as BAM3 and J774A.1, a fibroblast cell line NIH3T3, and thymocytes. MFG-E8 has

a signal sequence at the amino-terminus, but has no putative membrane-spanning region, suggesting that it is a secreted protein. In fact, the culture supernatant of thioglycollate-elicited peritoneal macrophages contained a large amount of MFG-E8 of M_r 74K (Fig. 2d). On the other hand, P388D1 cells secreted only negligible levels of MFG-E8, although the cell lysates contained a substantial amount of this compound.

Recombinant MFG-E8-L was produced in human 293T cells, and purified. MFG-E8-L did not bind freshly isolated thymocytes. On the other hand, when thymocytes were treated with dexamethasone, MFG-E8-L bound the apoptotic cells, as revealed by double staining with MFG-E8-L and TUNEL (Fig. 3b). MFG-E8-L contains two EGF (epidermal growth factor) domains, a proline/threonine-rich domain (P/T-rich domain), and two factor-VIII-homologous domains (C1 and C2; Fig. 3a). MFG-E8-S is coded for an alternatively spliced form of MFG-E8 mRNA, and lacks the P/T-rich domain. An analysis by fluorescence-activated cell sorting (FACS) indicated that not only MFG-E8-L, but also MFG-E8-S, D89E, which carried a point mutation in the RGD motif, and C1C2, which contained only the C1 and C2 domains, bound to apoptotic thymocytes (Supplementary Information Fig. 1). Annexin V binds apoptotic cells by recognizing phosphatidylserine (PS)⁷. When apoptotic thymocytes were pretreated with MFG-E8-L, the binding of annexin V to apoptotic cells was severely inhibited, whereas the preincubation of apoptotic thymocytes with annexin V inhibited binding of MFG-E8-L to apoptotic cells (Fig. 3c), suggesting that MFG-E8-L bound to PS. To directly examine this possibility, microtitre plates were coated with various phospholipids, and incubated with MFG-E8-L. As shown in Fig. 3d, MFG-E8-L bound to plates coated with PS or with phosphatidylethano-

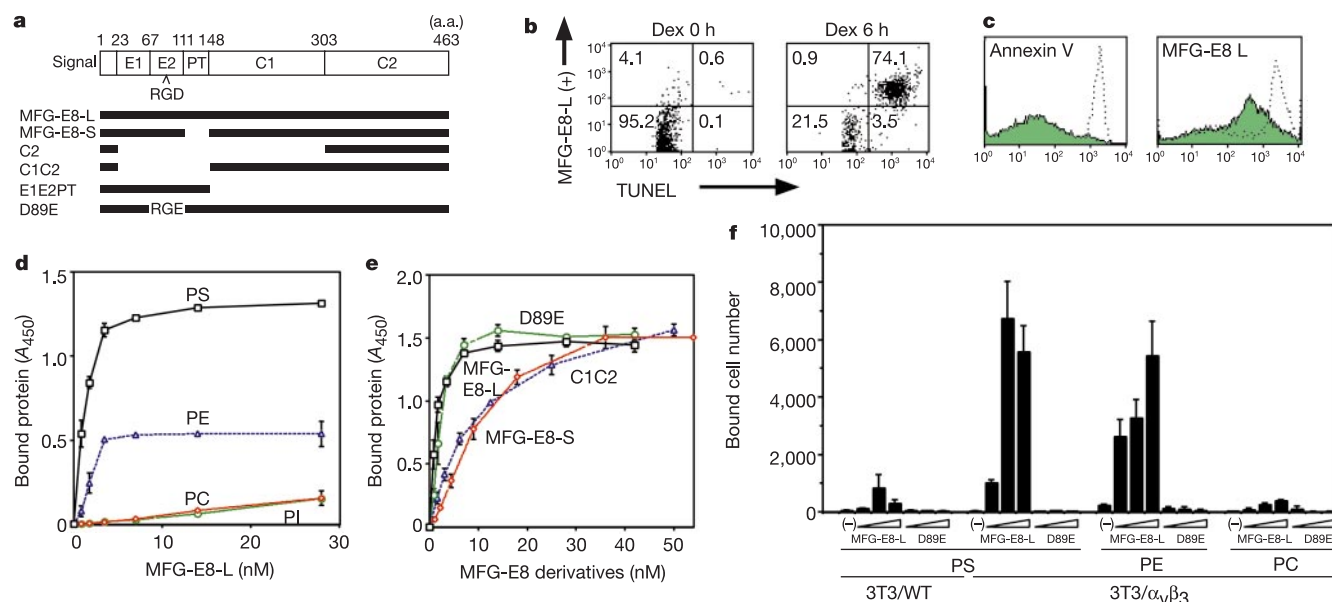


Figure 3 Binding of MFG-E8 to aminophospholipids and integrin-expressing cells. **a**, The structure of MFG-E8-L is shown at the top. MFG-E8-S lacks the P/T-rich domain. In C2 and C1C2, the signal sequence was fused to the C2 or C1-C2 domain. E1E2PT is a truncated form lacking the C1 and C2 domains. In D89E, the aspartic acid in the RGD motif was replaced by glutamic acid. a.a., amino acid. **b**, Thymocytes not treated or treated with dexamethasone for 6 h were incubated with MFG-E8-L, and stained with biotinylated 2422 antibody and PE-streptavidin. The cells were then stained with TUNEL, and analysed by FACS. **c**, Left panel, thymocytes treated with dexamethasone for 6 h were preincubated with 1.25 $\mu\text{g ml}^{-1}$ MFG-E8-L, and stained with 10 $\mu\text{g ml}^{-1}$ PE-annexin V. Right panel, the apoptotic thymocytes were preincubated with 10 $\mu\text{g ml}^{-1}$ of annexin V, and stained with 0.25 $\mu\text{g ml}^{-1}$ MFG-E8-L. In each panel, the annexin V- and MFG-E8-L-

staining profile without competitor is shown by dotted lines. **d**, Microtitre plates coated with PS, PE, PC or PI were incubated with MFG-E8-L, and MFG-E8-L bound to the wells was quantified by ELISA. Assays were done in triplicate, and the average values are plotted with s.d. A_{450} , absorbance at 450nm. **e**, Microtitre plates coated with PS were incubated with the indicated MFG-E8 derivatives, and MFG-E8 bound to the wells was quantified by ELISA. **f**, Microtitre wells coated with PS, PE or PC were successively incubated with three different concentrations (0.1, 1.0 and 2.0 $\mu\text{g ml}^{-1}$) of MFG-E8-L or D89E, and with NIH3T3 (3T3/WT) or $\alpha_v\beta_3$ integrin-expressing transformants (3T3/ $\alpha_v\beta_3$). The number of cells attached to the wells was quantified as described in Methods. The assays were done three times, and average values are plotted. WT, wild type.

lamine (PE), but not to the plates coated with phosphatidylcholine (PC) or phosphatidylinositol (PI). D89E bound to the PS-coated plates as efficiently as did the wild-type MFG-E8-L (Fig. 3e). On the other hand, the affinity of MFG-E8-S and C1C2 for the PS-coated plate was about eight times lower than that of MFG-E8-L. These results indicate that MFG-E8-L could recognize aminophospholipids via its C1C2 domain, and that the P/T-rich domain present in MFG-E8-L contributed to MFG-E8-L's affinity for these phospholipids.

The second EGF domain of MFG-E8 contains an RGD motif that could be recognized by some members of the integrin family⁸. We therefore considered the possibility that MFG-E8-L works as a bridge between apoptotic cells expressing aminophospholipids and phagocytes expressing integrins. Mouse NIH3T3 cell transformants that express a high level of $\alpha_v\beta_3$ integrin were established (Supplementary Information Fig. 2). MFG-E8-L did not specifically bind to NIH3T3 or its $\alpha_v\beta_3$ integrin transformants (data not shown). On the other hand, MFG-E8 enhanced adherence of

NIH3T3 to the PS-coated plates (Fig. 3f). The D89E mutant was incapable of mediating the adhesion of NIH3T3 to the PS-coated wells, indicating that this effect of MFG-E8-L was dependent on its RGD motif. When NIH3T3 expressing $\alpha_v\beta_3$ integrin was used as the target, the effect of MFG-E8-L on the adhesion of cells to the PS-coated wells was more dramatic. That is, 20 cells adhered to the untreated wells or the wells treated with D89E, whereas about 7,000 cells adhered to the wells that were pretreated with $1.0 \mu\text{g ml}^{-1}$ MFG-E8-L. MFG-E8-L also supported the adherence of NIH3T3 to PE-coated wells, but not to the PC-coated wells (Fig. 3f).

We then examined whether MFG-E8-L could stimulate NIH3T3 to engulf apoptotic cells. When freshly prepared thymocytes from ICAD-Sdm mice were co-cultured with NIH3T3, no thymocytes were attached to or engulfed by NIH3T3 (Fig. 4a). On the other hand, when NIH3T3 cells were co-cultured with apoptotic thymocytes, about 6% of the NIH3T3 cells engulfed more than 4 thymocytes each. The presence of MFG-E8-L increased the percent-

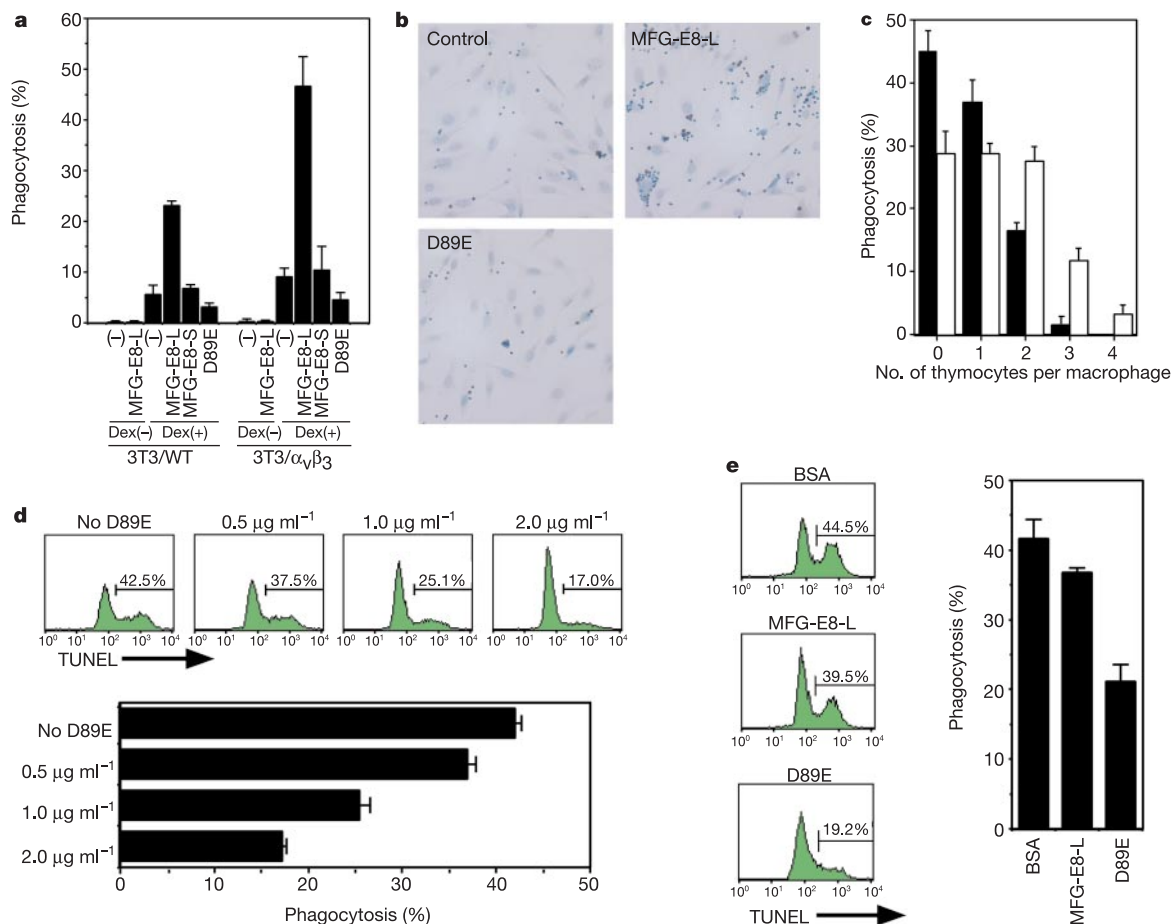


Figure 4 MFG-E8-L-dependent engulfment of apoptotic cells. **a**, Thymocytes from ICAD-Sdm mice were not treated or treated for 4 h with dexamethasone. NIH3T3 (3T3/WT) or its $\alpha_v\beta_3$ integrin transformant (3T3/ $\alpha_v\beta_3$) was co-cultured with the thymocytes in the absence (–) or presence (+) of $0.1 \mu\text{g ml}^{-1}$ of MFG-E8-L, MFG-E8-S, or D89E, and the percentage of the cells that engulfed more than 4 thymocytes was determined. The experiments were performed twice in triplicate, and the average number is shown with s.d. **b**, The NIH3T3 transformants expressing $\alpha_v\beta_3$ integrin were incubated with apoptotic thymocytes in the absence (control) or presence of MFG-E8-L or D89E, and were observed by light microscopy. Magnification, $\times 100$. **c**, Phagocytosis of apoptotic thymocytes was carried out with the resident peritoneal macrophages in the absence (closed bars) or presence of $0.1 \mu\text{g ml}^{-1}$ of MFG-E8-L (open bars). The percentages of

macrophages that carried none, 1, 2, 3 or ≥ 4 TUNEL-positive cells were determined. The experiments were done three times. **d**, Thymocytes from ICAD-Sdm mice were treated for 4 h with dexamethasone, and co-cultured with thioglycollate-elicited peritoneal macrophages in the presence of the indicated concentrations of D89E. The cells were then stained for Mac-1 and TUNEL. The FACS profile for TUNEL-positive cells in the Mac-1⁺ cell population is shown with the percentages of TUNEL-positive macrophages. The experiments were carried out three times, and the average values are plotted with s.d. in the lower panel. **e**, Mice pretreated with thioglycollate were injected i.p. with apoptotic thymocytes with $10 \mu\text{g}$ of BSA, MFG-E8-L or D89E. After 90 min, macrophages were collected, and stained for Mac-1 and TUNEL. The experiments were done 5 times, and the average values are plotted in the right panel.

tage of NIH3T3 cells carrying more than 4 thymocytes inside to 23%. The effect of MFG-E8-L on phagocytosis was more pronounced when the $\alpha_v\beta_3$ integrin transformants were used as phagocytes. In this case, the percentage of the cells that engulfed more than 4 thymocytes increased from 9% to 46% by MFG-E8-L, and about 20% of the cells carried more than 7 thymocytes inside (Fig. 4b). There was an optimal concentration of MFG-E8-L to enhance the phagocytosis (Supplementary Information Fig. 3). Below $0.1 \mu\text{g ml}^{-1}$, MFG-E8-L enhanced the phagocytosis, whereas it had an inhibitory effect above this concentration. This inhibition was relieved by adding the 2422 antibody. A similar effect was observed with the resident peritoneal macrophages. That is, the percentage of macrophages that engulfed thymocytes after co-culture with apoptotic thymocytes increased from 55% to 70%, and the percentages of the macrophages engulfing more than 3 thymocytes increased from 1.6% to 15.1% (Fig. 4c).

D89E inhibited the phagocytic activity of NIH3T3 and its transformants (Fig. 4a and b), suggesting that it works as a dominant-negative form of MFG-E8. To assess the involvement of MFG-E8-L on the phagocytosis of apoptotic cells by thioglycollate-elicited peritoneal macrophages, we used this property of D89E. As shown in Fig. 4d, when thioglycollate-elicited peritoneal macrophages were co-cultured with apoptotic thymocytes from ICAD-Sdm mice, approximately 42% of the macrophages became TUNEL-positive. The appearance of TUNEL-positive cells, and thus the phagocytosis of thymocytes by macrophages, was inhibited by D89E in a dose-dependent manner. This inhibitory effect of D89E was also observed *in vivo*. Sterile peritonitis was induced in mice by administration of thioglycollate (Fig. 4e). Four days later, the mice received an intraperitoneal (i.p.) injection of apoptotic thymocytes from ICAD-Sdm mice. After 90 min, about 45% of Mac-1⁺ macrophages in the peritoneal cavity became TUNEL-positive. This appearance of TUNEL-positive macrophages was greatly inhibited by co-injection with D89E, but less significantly by co-injection with the wild-type MFG-E8-L. These results indicated that MFG-E8-L played a vital role in the phagocytosis of apoptotic cells by thioglycollate-elicited peritoneal macrophages.

Many proteins have been proposed to be receptors for apoptotic cells or bridging proteins between apoptotic cells and phagocytes^{1,2,9–12}. However, it has been unclear whether these proteins directly bind to apoptotic cells, and (or) how these proteins are involved in phagocytosis. Here we have shown that MFG-E8-L specifically bound to apoptotic cells by recognizing aminophospholipids such as PS and PE. Aminophospholipids localized to the inner leaflet of the plasma membrane in healthy cells are exposed on the cell surface when cells are triggered to undergo apoptosis^{13–15}. Cells repleted with PS can be recognized by phagocytes as targets¹⁶. These results indicate that the exposed PS fulfils the criteria for an 'eat me' signal. Most of the molecules proposed to be receptors for apoptotic cells bind not only PS but also PI^{3,17}. On the other hand, MFG-E8-L exclusively bound PS and PE, supporting the idea that MFG-E8-L specifically recognizes apoptotic cells. Integrins have been suggested as receptors for apoptotic cells in several systems^{18,19}. However, because neither $\alpha_v\beta_3$ or $\alpha_v\beta_5$ integrin can bind PS, it has not been clear how these integrins recognize apoptotic cells. MFG-E8-L seems to resolve this dilemma. There are at least two phagocytosis systems in *Caenorhabditis elegans*²⁰. How mammalian phagocytes use this or other systems such as phosphatidylserine receptor (PSR)²¹ or a protooncogene expressed in monocytes and tissues of epithelial and reproductive origin (MER)²² remains to be studied. Finally, MFG-E8 was originally identified as one of the most abundant proteins in the membranes of milk fat globules³. Mammary glands undergo massive involution when suckling or milking ceases²³. During this process, a large number of epithelial cells are killed by apoptosis, and are cleared to ensure the remodelling of the gland²⁴. Identification of MFG-E8-L as a molecule that

recognizes the apoptotic cells would help to elucidate the molecular mechanism behind this involution and remodelling. □

Methods

Cell lines, antibodies and recombinant MFG-E8

Mouse NIH3T3 transformants expressing integrin $\alpha_v\beta_3$ were established by infecting cells with the retrovirus carrying mouse integrin α_v and β_3 cDNAs^{25,26}. To produce monoclonal antibodies, 1.5×10^7 thioglycollate-elicited peritoneal macrophages were injected subcutaneously into Armenian hamsters with a 4-week interval. Cells from the lymph nodes were fused with mouse myeloma, the supernatants of hybridomas were tested by a phagocytosis assay, and monoclonal antibodies were purified by protein A-Sepharose. Rabbit antibody against a peptide (CNSHKKNIFEKPFMAR) of mouse MFG-E8 was prepared at the Peptide Institute (Osaka, Japan). The antibody was affinity-purified using AF-amino-Toyopearl (Tosoh) to which the peptide had been attached. MFG-E8-L and its Flag-tagged derivatives were expressed in human 293T cells, and the proteins secreted into the medium was purified using 2422 antibody- or anti-Flag M2 affinity gel (Sigma).

Phagocytosis assay

Thymocytes from ICAD-Sdm mice⁵ were incubated at 37 °C with $10 \mu\text{M}$ dexamethasone in DMEM containing 10% FCS. C57BL/6 mice (12 weeks old) received an intraperitoneal (i.p.) injection of 3% (w/v) thioglycollate (Sigma). For *in vitro* phagocytosis assay, the elicited peritoneal macrophages were collected after 4 days, and cultured in DMEM containing 10% FCS. Thymocytes (1×10^6 cells) were added to 2.5×10^5 macrophages on 48-well plates. For *in vivo* analysis²⁷, the mice treated with thioglycollate for 4 days were injected i.p. with apoptotic thymocytes (9×10^7 cells). Phagocytosis was allowed to proceed for 1.5 h *in vitro* and *in vivo*. Macrophages were collected from the plates or peritoneal cavities, and incubated on ice for 30 min with $4 \mu\text{g ml}^{-1}$ phycoerythrin (PE)-conjugated rat anti-mouse Mac-1 (BD PharMingen) in FACS staining buffer (PBS containing 2% FCS and 0.02% NaN_3) in the presence of $2.5 \mu\text{g ml}^{-1}$ rat anti-mouse Fc γ III/II receptor (BD PharMingen). Cells were fixed with 1% paraformaldehyde, treated with 0.1% Triton X-100, and suspended in 100 μl of 100 mM cacodylate buffer (pH 7.2) containing 1 mM CoCl_2 and 0.01% BSA. The TUNEL reaction was carried out at 37 °C for 45 min with 100 units ml^{-1} terminal deoxynucleotidyl transferase and $2.5 \mu\text{M}$ FITC-labelled dUTP (Roche Diagnostics), and analysed by flow cytometry. Phagocytosis was also evaluated by observing the cells under a microscope. In brief, peritoneal macrophages (1×10^5 cells) or NIH3T3 cells (2×10^4 cells) were cultured in 8-well Lab-Tek II chamber slides (Nalge Nunc), and phagocytosis was allowed to proceed as described above. After fixation, the cells were subjected to the TUNEL reaction using an Apoptag kit (Intergen), and observed by light microscopy. The number of macrophages that carried TUNEL-positive thymocytes was counted, and the percentage of these cells per total number of cells (150 cells) was determined. Because the TUNEL-reactivity of the phagocytosed cells in NIH3T3 cells was low probably due to the low level of DNase II²⁸, we counted as positive those NIH3T3 cells carrying more than 4 thymocytes, to exclude the population of NIH3T3 which thymocytes were just bound to, but not engulfed by.

Identification of MFG-E8

The 2422 antibody was covalently linked to protein A-Sepharose (2 mg protein per ml bed volume) using dimethyl pimelimidate (Pierce). Mouse P388D1 cells (2.4×10^7 cells) were lysed in RIPA buffer (50 mM HEPES-NaOH buffer, pH 7.6, containing 1% Triton X-100, 0.1% SDS, 0.5% sodium deoxycholate, 150 mM NaCl, 1.5 mM MgCl_2 , 1 mM EGTA, 10% glycerol, 1 mM (*p*-aminodiphenyl) methanesulphonyl fluoride hydrochloride, $1 \mu\text{g ml}^{-1}$ leupeptin and $1 \mu\text{g ml}^{-1}$ pepstatin). The lysate was precleared with 3 ml of human IgG Sepharose, and incubated for 2 h with $150 \mu\text{l}$ of the 2422 antibody protein A-Sepharose. After washing with RIPA buffer containing 0.5 M NaCl, proteins bound to the beads were eluted with 100 mM triethylamine (pH 11.5) containing 0.1% Triton X-100, separated by electrophoresis on a 10% polyacrylamide gel, and blotted onto a polyvinylidene difluoride (PVDF) membrane. The immobilized protein was reduced, S-carboxymethylated, and digested with *Achromobacter* protease I as described²⁹. Peptides released from the membrane were analysed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.

Solid-phase ELISA and cell adhesion assay

The solid phase ELISA (enzyme-linked immunosorbent assay) for MFG-E8 binding to phospholipids was carried out as described³⁰. In brief, phospholipid solutions ($3 \mu\text{g ml}^{-1}$, 100 μl) in methanol were added to 96-well plates, and air-dried. The wells were treated with 10 mg ml^{-1} bovine serum albumin (BSA) in PBS. MFG-E8 was added to the wells, and incubated at room temperature for 1 h. After washing with PBS containing 0.05% Tween 20, MFG-E8 bound to the wells was quantified by ELISA with biotinylated anti-Flag antibody and peroxidase-conjugated streptavidin. To assay the ability of MFG-E8 to link the cells to phospholipids, MFG-E8 was bound to microtitre plates coated with phospholipids as described above. Cells (4×10^4) in Tyrode buffer (5 mM HEPES-NaOH buffer, pH 7.4, 135 mM NaCl, 5.4 mM KCl, 1.0 mM MgCl_2 , 10 mM glucose and 10 mg ml^{-1} BSA) was added to each well, and incubated at room temperature for 1 h. The cells that had adhered to the plates were quantified by a CyQUANT Cell Proliferation Assay kit (Molecular Probes) using a fluorescent microplate reader (BioLumin 960, Molecular Dynamics) set at an excitation wavelength of 485 nm and an emission wavelength of 520 nm.

Received 11 January; accepted 20 February 2002.

- Savill, J. & Fadok, V. Corpse clearance defines the meaning of cell death. *Nature* **407**, 784–788 (2000).
- Platt, N., da Silva, R. P. & Gordon, S. Recognizing death: the phagocytosis of apoptotic cells. *Trends Cell Biol.* **8**, 365–372 (1998).
- Stubbs, J. D. *et al.* cDNA cloning of a mouse mammary epithelial cell surface protein reveals the existence of epidermal growth factor-like domains linked to factor VIII-like sequences. *Proc. Natl Acad. Sci. USA* **87**, 8417–8421 (1990).
- Oshima, K. *et al.* Lactation-dependent expression of an mRNA splice variant with an exon for a multiply O-glycosylated domain of mouse milk fat globule glycoprotein MFG-E8. *Biochem. Biophys. Res. Commun.* **254**, 522–528 (1999).
- McIlroy, D. *et al.* An auxiliary mode of apoptotic DNA fragmentation provided by phagocytes. *Genes Dev.* **14**, 549–558 (2000).
- Bowman, E. J., Siebers, A. & Altendorf, K. Bafilomycins: a class of inhibitors of membrane ATPases from microorganisms, animal cells, and plant cells. *Proc. Natl Acad. Sci. USA* **85**, 7972–7976 (1988).
- Koopman, G. *et al.* Annexin V for flow cytometric detection of phosphatidylserine expression on B cells undergoing apoptosis. *Blood* **84**, 1415–1420 (1994).
- Hynes, R. O. Integrins: versatility, modulation, and signalling in cell adhesion. *Cell* **69**, 11–25 (1992).
- Fadok, V. A., Bratton, D. L., Frasch, S. C., Warner, M. L. & Henson, P. M. The role of phosphatidylserine in recognition of apoptotic cells by phagocytes. *Cell Death Differ.* **5**, 551–562 (1998).
- Nakano, T. *et al.* Cell adhesion to phosphatidylserine mediated by a product of growth arrest-specific gene 6. *J. Biol. Chem.* **272**, 29411–29414 (1997).
- Savill, J., Hogg, N., Ren, Y. & Haslett, C. Thrombospondin cooperates with CD36 and the vitronectin receptor in macrophage recognition of neutrophils undergoing apoptosis. *J. Clin. Invest.* **90**, 1513–1522 (1992).
- Ogden, C. A. *et al.* C1q and mannose binding lectin engagement of cell surface calreticulin and CD91 initiates macropinocytosis and uptake of apoptotic cells. *J. Exp. Med.* **194**, 781–795 (2001).
- Fadok, V. A. *et al.* Different populations of macrophages use either the vitronectin receptor or the phosphatidylserine receptor to recognize and remove apoptotic cells. *J. Immunol.* **149**, 4029–4035 (1992).
- Emoto, K., Toyama-Sorimachi, N., Karasuyama, H., Inoue, K. & Umeda, M. Exposure of phosphatidylethanolamine on the surface of apoptotic cells. *Exp. Cell Res.* **232**, 430–434 (1997).
- Blankenberg, F. G. *et al.* In vivo detection and imaging of phosphatidylserine expression during programmed cell death. *Proc. Natl Acad. Sci. USA* **95**, 6349–6354 (1998).
- Fadok, V. A., de Cathelineau, A., Daleke, D. L., Henson, P. M. & Bratton, D. L. Loss of phospholipid asymmetry and surface exposure of phosphatidylserine is required for phagocytosis of apoptotic cells by macrophages and fibroblasts. *J. Biol. Chem.* **276**, 1071–1077 (2001).
- Rigotti, A., Acton, S. L. & Krieger, M. The class B scavenger receptors SR-BI and CD36 are receptors for anionic phospholipids. *J. Biol. Chem.* **270**, 16221–16224 (1995).
- Savill, J., Dransfield, I., Hogg, N. & Haslett, C. Vitronectin receptor-mediated phagocytosis of cells undergoing apoptosis. *Nature* **343**, 170–173 (1990).
- Albert, M. L., Sauter, B. & Bhardwaj, N. Dendritic cells acquire antigen from apoptotic cells and induce class I-restricted CTLs. *Nature* **392**, 86–89 (1998).
- Hengartner, M. O. Apoptosis: corralling the corpses. *Cell* **104**, 325–328 (2001).
- Fadok, V. A. *et al.* A receptor for phosphatidylserine-specific clearance of apoptotic cells. *Nature* **405**, 85–90 (2000).
- Scott, R. S. *et al.* Phagocytosis and clearance of apoptotic cells is mediated by MER. *Nature* **411**, 207–211 (2001).
- Wilde, C. J., Knight, C. H. & Flint, D. J. Control of milk secretion and apoptosis during mammary involution. *J. Mamm. Gland Biol. Neoplasia* **4**, 129–136 (1999).
- Fadok, V. A. Clearance: the last and often forgotten stage of apoptosis. *J. Mamm. Gland Biol. Neoplasia* **4**, 203–211 (1999).
- Wada, J. *et al.* Cloning of mouse integrin α_V cDNA and role of the α_V -related matrix receptors in metanephric development. *J. Cell Biol.* **132**, 1161–1176 (1996).
- McHugh, K. P., Kitazawa, S., Teitelbaum, S. L. & Ross, F. P. Cloning and characterization of the murine $\beta(3)$ integrin gene promoter: identification of an interleukin-4 responsive element and regulation by STAT-6. *J. Cell Biochem.* **81**, 320–332 (2001).
- Taylor, P. R. *et al.* A hierarchical role for classical pathway complement proteins in the clearance of apoptotic cells in vivo. *J. Exp. Med.* **192**, 359–366 (2000).
- Kawane, K. *et al.* Requirement of DNase II for definitive erythropoiesis in the mouse fetal liver. *Science* **292**, 1546–1549 (2001).
- Iwamatsu, A. & Yoshida-Kubomura, N. Systematic peptide fragmentation of polyvinylidene difluoride (PVDF)-immobilized proteins prior to microsequencing. *J. Biochem. (Tokyo)* **120**, 29–34 (1996).
- Andersen, M. H., Berglund, L., Rasmussen, J. T. & Petersen, T. E. Bovine PAS-6/7 binds $\alpha_5\beta_3$ integrins and anionic phospholipids through two domains. *Biochemistry* **36**, 5441–5446 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank T. Kitamura for the retrovirus expression system, Y. Seto for maintaining the mice, and S. Aoyama for secretarial assistance. This work was supported in part by Grants-in-Aid from the Ministry of Education, Science, Sports and Culture in Japan.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S. N. (e-mail: nagata@genetic.med.osaka-u.ac.jp).

Dissection of COPI and Arf1 dynamics *in vivo* and role in Golgi membrane transport

John F. Presley[†], Theresa H. Ward^{*}, Andrea C. Pfeifer^{*}, Eric D. Siggia[‡], Robert D. Phair[§] & Jennifer Lippincott-Schwartz^{*}

^{*} Cell Biology and Metabolism Branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA

[‡] Center for Studies in Physics and Biology, The Rockefeller University, New York, New York 10021, USA

[§] Bioinformatics Services, Rockville, Maryland 20854, USA

Cytosolic coat proteins that bind reversibly to membranes have a central function in membrane transport within the secretory pathway^{1,2}. One well-studied example is COPI or coatomer, a heptameric protein complex that is recruited to membranes by the GTP-binding protein Arf1. Assembly into an electron-dense coat then helps in budding off membrane to be transported between the endoplasmic reticulum (ER) and Golgi apparatus². Here we propose and corroborate a simple model for coatomer and Arf1 activity based on results analysing the distribution and lifetime of fluorescently labelled coatomer and Arf1 on Golgi membranes of living cells. We find that activated Arf1 brings coatomer to membranes. However, once associated with membranes, Arf1 and coatomer have different residence times: coatomer remains on membranes after Arf1-GTP has been hydrolysed and dissociated. Rapid membrane binding and dissociation of coatomer and Arf1 occur stochastically, even without vesicle budding. We propose that this continuous activity of coatomer and Arf1 generates kinetically stable membrane domains that are connected to the formation of COPI-containing transport intermediates. This role for Arf1/coatomer might provide a model for investigating the behaviour of other coat protein systems within cells.

To study coatomer dynamics *in vivo*, variants of green fluorescent protein (GFP), namely enhanced GFP (EGFP) and enhanced yellow fluorescent protein (EYFP), were fused to the carboxy terminus of the ϵ COP subunit of coatomer (ϵ COP-GFP or ϵ COP-YFP) and expressed in *ldlf* cells, which contain a mutated ϵ COP that at 40 °C is degraded, causing coatomer inactivity and growth inhibition³. The cells grew indefinitely at 40 °C (Fig. 1a), indicating that ϵ COP-GFP could substitute functionally for endogenous ϵ COP in coatomer complexes in these cells. Subsequent experiments were performed with stable *ldlf* cell lines expressing ϵ COP-GFP (or ϵ COP-YFP) that were selected by continuous growth at 40 °C unless indicated otherwise. ϵ COP-GFP in these cells localized in a manner that was indistinguishable from that of endogenous coatomer detected by antibody staining of β COP (Fig. 1b). Biochemical (Supplementary Information) and kinetic (see below) analyses revealed that most ϵ COP-GFP molecules were stably assembled into coatomer complexes that underwent binding and dissociation from membranes.

The identity and behaviour of coatomer-containing membranes, assessed by ϵ COP-YFP labelling, were studied in dual-colour time-lapse experiments in cells co-expressing cyan fluorescent protein (CFP)-tagged secretory cargo markers. ϵ COP-YFP was present on juxtanuclear Golgi membranes, as well as on ER-to-Golgi (anterograde) transport intermediates containing vesicular stomatitis virus ts045 G protein (VSVG) tagged with CFP in cells shifted from

[†] Present address: Department of Anatomy and Cell Biology, McGill University, Montreal, Quebec, Canada.

40 °C to 15 °C (Fig. 1c)⁴. On being warmed to 32 °C, the anterograde intermediates moved as globular or tubular shapes towards the Golgi complex, with ϵ COP–YFP remaining localized on their surfaces (Fig. 1d, e). ϵ COP–YFP was only weakly associated with presumptive retrograde, Golgi-to-ER, transport intermediates, which appeared as Golgi tubule extensions when immunolabelled with the ERGIC53 homologue p58 (not shown), an ER/Golgi cycling molecule⁵, or when co-stained with p58–CFP (Fig. 1f). No small fluorescent objects with the predicted size of vesicles (that is, one pixel) were observed moving out from any of these structures (Supplementary Information). These findings are consistent with previous reports of coatomer distribution within living cells^{6,7}.

To measure how rapidly coatomer dissociates from Golgi membranes, we performed fluorescence loss in photobleaching (FLIP) experiments⁸. On repeated bleaching outside the Golgi, all cellular ϵ COP–GFP fluorescence was lost in a few minutes (Fig. 2a), indicating that coatomer resides on Golgi membranes only transiently before dissociating and exchanging with cytoplasmic pools. To determine the rate at which coatomer rebinds to Golgi membranes, we selectively photobleached the Golgi and then watched for

fluorescence recovery after photobleaching (FRAP)⁸. The prebleach, Golgi-to-cytoplasmic fluorescence ratio (that is, 40:60) recovered exponentially with a half-time ($t_{1/2}$) of 35 s (Fig. 2b, c) and was not affected by microtubule disruption (which inhibits the movement of anterograde transport intermediates) (not shown). This indicated that recovery was mediated not by delivery via anterograde intermediates but by exchange between Golgi-bound and freely mobile cytoplasmic pools of coatomer. FRAP experiments examining ϵ COP–GFP dynamics on anterograde intermediates revealed that ϵ COP–GFP also rapidly exchanged on and off these structures (Fig. 2d), whether or not they were moving, and at a rate similar to that at the Golgi complex. In cells co-expressing the active GTP-bound form of Arf1 [Q71L] that is unable to hydrolyse GTP⁹, or in cells treated with GTP- γ S (not shown), ϵ COP–GFP became irreversibly bound to membranes and no fluorescence recovery occurred during FRAP (Fig. 2c, e). These results indicate that coatomer dissociation from membranes *in vivo* occurs rapidly and is dependent on Arf1–GTP hydrolysis¹⁰.

To determine whether coatomer dissociation from membranes coincides with Arf1–GTP hydrolysis, we examined by FRAP the

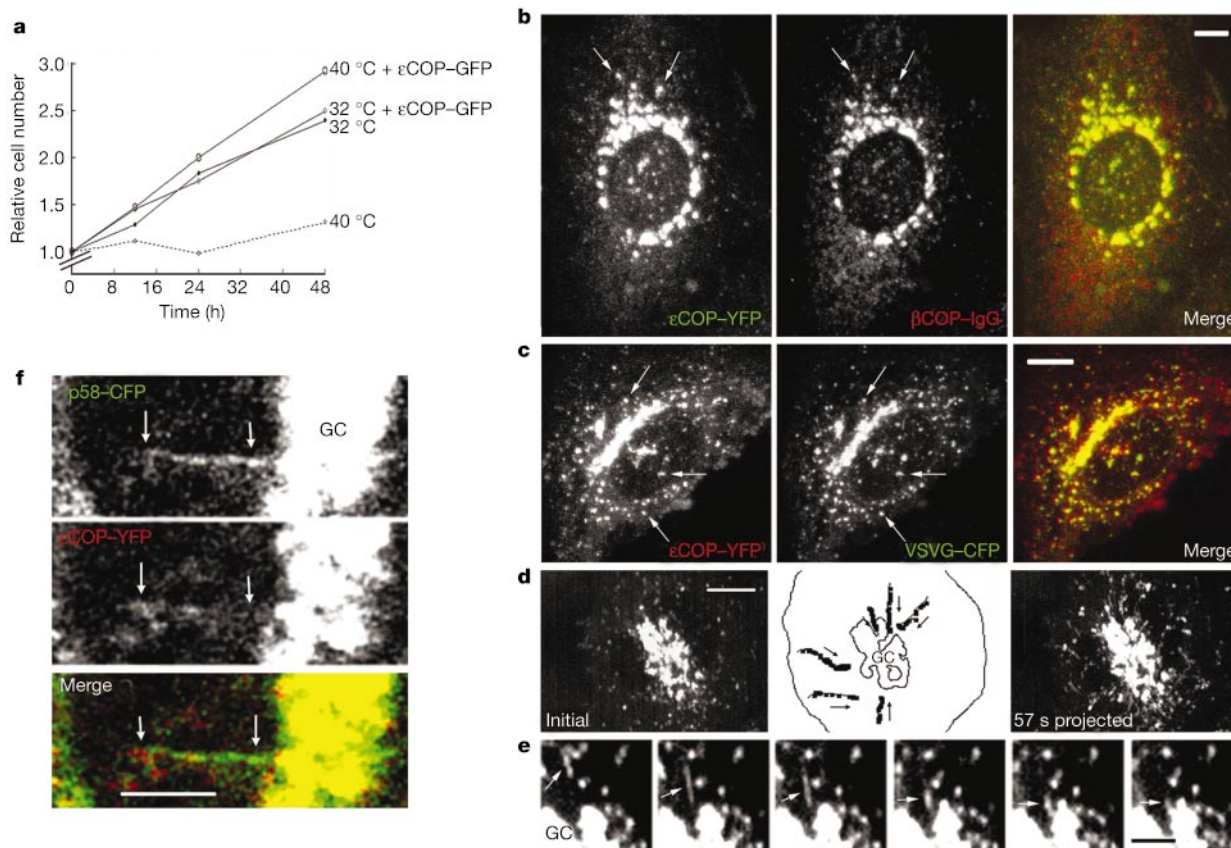


Figure 1 Incorporation of ϵ COP–GFP into functional coatomer complexes and their distribution. **a**, Growth rate (cell number relative to initial cell number with time) at 32 or 40 °C in untreated *IdIF* cells and *IdIF* cells expressing ϵ COP–GFP by transient transfection. **b**, Co-localization of ϵ COP–GFP (left) and β COP stained with anti- β COP antibody (generated from the EAGE peptide) single-letter amino-acid codes) in β COP) and rhodamine secondary antibody (middle) in an *IdIF* cell stably expressing ϵ COP–GFP. Merged image (right) shows co-localization (yellow) of the markers in Golgi and peripheral structures (see arrows). Scale bar, 5 μ M. **c**, *IdIF* cells stably expressing ϵ COP–YFP were transfected with VSVG–CFP and incubated at 40 °C to accumulate VSVG–CFP in the ER. The temperature was lowered to 15 °C for 3 h to accumulate VSVG–CFP in ER-to-Golgi transport intermediates. Extensive co-localization (yellow, merged image) of the two markers was found in peripheral (arrowed) and juxtanuclear structures. Scale bar, 5 μ M.

d, Inward movement of peripheral ϵ COP–GFP-containing structures in *IdIF* cells stably expressing ϵ COP–GFP. Shown are the initial image of the time series (left), several representative tracks taken by peripheral structures during the time series (middle), and a maximum-value projection of the entire time series (acquired at two images s^{-1} for 57 s) showing paths taken by translocating structures (right). Scale bar, 5 μ M. (See Quicktime movie in Supplementary Information.) **e**, Area near the Golgi complex (GC) from the time series in **d** illustrating tubulation of a ϵ COP–GFP labelled structure as it translocates towards the Golgi complex. Scale bar, 2 μ M. **f**, Still frame from a time sequence in an NRK cell expressing p58–CFP (top) and ϵ COP–YFP (middle). The merged image (bottom) shows only low levels of ϵ COP–YFP on p58–CFP-containing Golgi tubules (arrowed). Scale bar, 5 μ M.

kinetics of Arf1 membrane association and dissociation (which is mediated by GTP binding and hydrolysis¹⁰) by using a GFP-tagged Arf1 (ref. 11) (Arf1-GFP). Addition of the GFP tag did not alter any of Arf1's essential roles: Arf1-GFP functionally replaced the endogenous copy of *ARF1* in a $\Delta arf2$ yeast strain, and when expressed as a GTP-locked form (that is, Arf1[Q71L]-GFP) it could protect endogenous COPI from membrane release in the presence of brefeldin A (BFA) (Supplementary Information, Figs S1 and S2). During FRAP, exchange of bleached Arf1-GFP in the Golgi with unbleached Arf1-GFP in the cytoplasm occurred rapidly (Fig. 3a), was not limited by cytoplasmic diffusion of Arf1-GFP and was dependent on GTP hydrolysis because exchange did not occur in cells injected with GTP- γ S (not shown). Recovery was faster than that reported previously for Arf1-GFP¹¹ because of differences in FRAP conditions (Supplementary Information). Because the $t_{1/2}$ for recovery of Arf1-GFP (15 s) was significantly faster than for coatomer (35 s) measured under similar conditions, the data indicated that coatomer is stabilized on membranes for additional periods after Arf1 undergoes GTP hydrolysis and dissociation.

We visualized Arf1 and coatomer dissociation from membranes directly in cells co-expressing Arf1-CFP and ϵ COP-YFP treated with BFA¹⁰. BFA blocks the binding of Arf1 to membranes and the recruitment of coatomer by Arf1 (ref. 12), allowing the dissociation of Arf1 and coatomer to be observed in the absence of rebinding. On treatment with BFA, Arf1-CFP and ϵ COP-YFP rapidly dissociated from Golgi membranes before Golgi disassembly (Fig. 3b). The half-time for dissociation of Arf1-CFP (13 s) was significantly faster than that for ϵ COP-YFP (30 s). Similar half-times were measured in cells expressing the chimaeras individually, and with Arf1 tagged with the small haemagglutinin epitope (HA-Arf1), a widely used Arf1 reporter¹³ (Supplementary Information, Fig. S3), supporting the conclusion that Arf1 and coatomer dissociate independently from Golgi membranes.

We used FRAP or treatment with BFA to examine the effect of coatomer on the rate of dissociation of Arf1-GFP from the Golgi^{14,15} in two situations: when coatomer is irreversibly membrane-bound by treatment with AIF (ref. 16), an activator of trimeric G proteins¹⁷; and after the depletion of coatomer from cells. Treatment with a range of AIF concentrations had no major effect on the FRAP kinetics of Arf1-GFP (Fig. 3f, g) despite irreversible binding of ϵ COP-GFP to membranes (Fig. 3d, e) (see also Supplementary Information, Fig. S4). Moreover, coatomer depletion (Fig. 4a) had no effect on the FRAP recovery rate of Arf1-GFP (not shown) or on the release rate of Arf1-GFP from the Golgi after treatment with BFA (Fig. 4b). Thus, whether coatomer is stabilized on or depleted from membranes has no detectable effect on the GAP-mediated, GTP-hydrolysis-dependent membrane release of Arf1-GFP *in vivo*.

On the basis of the above results and previous data we formulated a simple model of activity and turnover of Arf1 and coatomer on membrane (Fig. 4c). The model explicitly uncouples Arf1 and coatomer release from membranes and posits no feedback from coatomer to Arf1. It encompasses the following six processes: (1) binding of cytoplasmic Arf1 to Golgi membranes and activation by guanine nucleotide exchange^{18,19}, which is inhibited by BFA¹²; (2) recruitment of cytoplasmic coatomer to Golgi membranes by binding to activated Arf1 (ref. 20); (3) interaction of activated Arf1 with alternative effector molecules in Golgi membranes²¹; (4) GTPase-activating protein (GAP)-mediated hydrolysis of GTP bound to Arf1 (ref. 22) in either coatomer¹⁴ or alternative effector complexes²¹ with the concomitant release of Arf1-GDP to the cytoplasm, while coatomer remains Golgi-associated; (5) binding of coatomer to membrane cargo, soluble-cargo receptors or other Golgi proteins^{23,24}; and (6) uncoating or release of coatomer from Golgi membranes to cytoplasm, which is inhibited by AIF (ref. 16).

To test this model quantitatively we formulated it on the basis of standard principles of chemical kinetics, with each process contributing terms to a system of differential equations (Supplementary

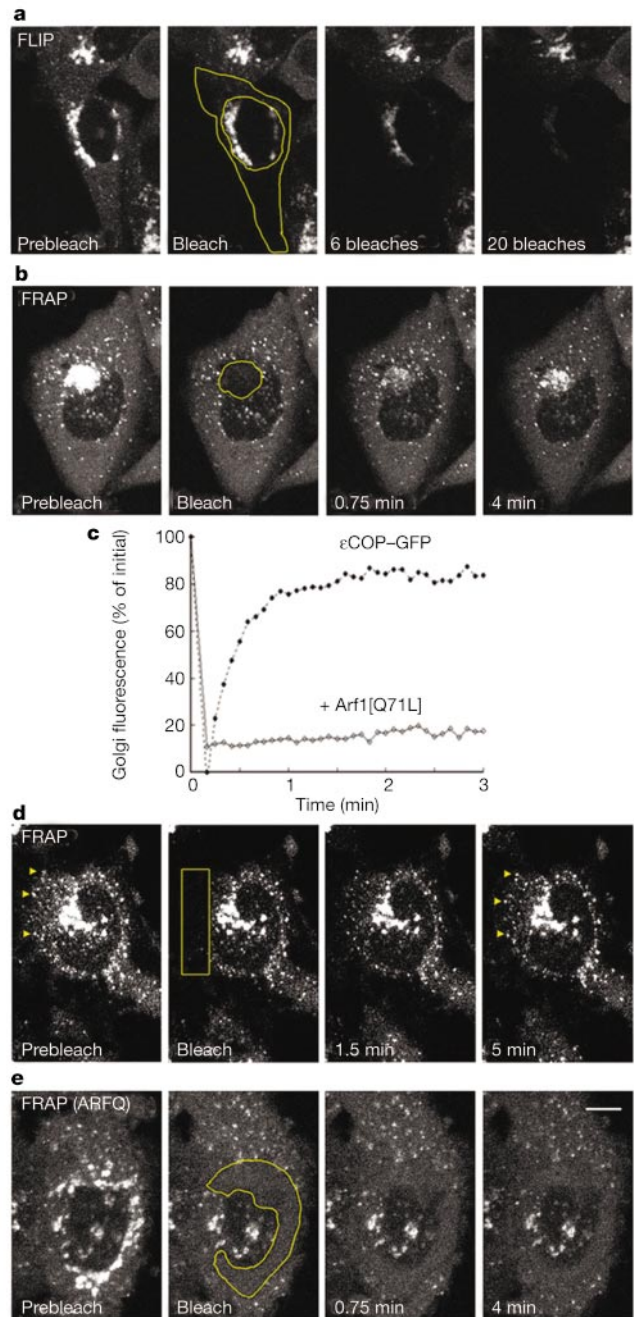


Figure 2 Kinetics of COPI binding to and dissociation from Golgi and ER-to-Golgi transport intermediates in *IdIF* cells stably expressing ϵ COP-GFP. **a**, Repeated photobleaching (FLIP) (one bleach every 15 s) of the cytoplasm (defined by the area between the two yellow lines). **b**, Golgi-FRAP. An initial prebleach image was taken (prebleach), then the Golgi region of interest (indicated by yellow outline) was bleached with high-intensity laser light. After the bleach, images were taken at 5–10-s intervals to monitor exchange between photobleached and non-bleached ϵ COP-GFP. Scale bar, 5 μ M. (See Quicktime movies in Supplementary Information.) **c**, Quantification of Golgi FRAP protocol with and without co-expression of Arf1[Q71L]. **d**, FRAP of ER-to-Golgi intermediates. To prevent movement of the intermediates, cells were chilled on ice for 20 min and warmed to 37 °C in the presence of 1 μ g ml⁻¹ nocodazole to depolymerize microtubules. The area within the yellow box in the cell periphery was illuminated with high-intensity laser light to photobleach the fluorescent structures containing ϵ COP-GFP. Arrowheads point to ER-to-Golgi intermediates that recovered their fluorescence while remaining at the same location. (See Quicktime movies in Supplementary Information.) **e**, FRAP of an area containing the Golgi region of interest (indicated by yellow outline) in a cell microinjected 5 h previously with a plasmid coding for the GTP-bound Arf1[Q71L] mutant (ARFQ). Golgi-associated fluorescence does not recover after photobleaching, indicating that exchange between Golgi and cytoplasmic ϵ COP-GFP is inhibited in these cells.

Information). We then examined whether it could account for the Arf1-GFP and ϵ COP-GFP photobleaching data (Figs 2c and 3a) and the Arf1-CFP and ϵ COP-YFP kinetics after treatment with BFA (Fig. 3b, c) while accounting for the steady-state abundances of Arf1 and coatamer on membranes and in cytoplasm. The latter was assessed by quantification of the fluorescent images with GFP standard solutions and by western blot analysis (Supplementary

Information). Significantly, we found that the model could account for all eight data sets by using the same set of parameter values (that is, rate constants). This is illustrated in Fig. 4d, by plotting the model solutions (smooth curves) on the same axes as the corresponding experimental data (dotted curves). The relevant rate constants were 0.0127 s^{-1} for binding of cytoplasmic Arf1 to Golgi (process 1), $1.05 \times 10^{-8} (\text{molecules per cell})^{-1} \text{ s}^{-1}$ for binding of cytoplasmic

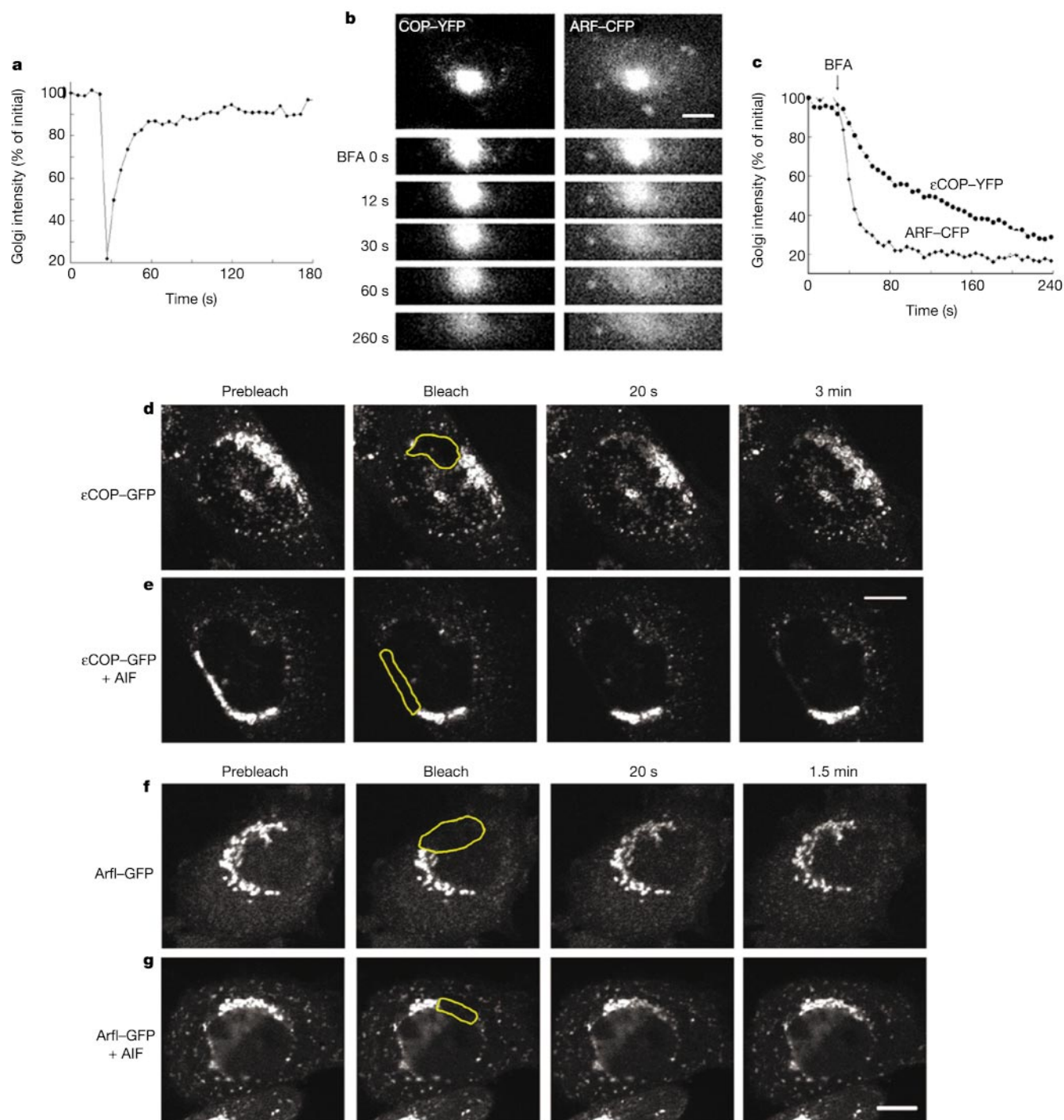


Figure 3 Golgi association-dissociation kinetics of Arf1-GFP/CFP and ϵ COP-GFP/YFP in AIF-treated or BFA-treated cells. **a**, Quantification of Golgi FRAP in CHO cell expressing Arf1-GFP. **b**, Kinetics of dissociation of ϵ COP-YFP and Arf1-CFP from Golgi membranes after the addition of BFA ($5 \mu\text{g ml}^{-1}$) in an *IdIF* cell co-expressing both proteins. **c**, Quantification of dissociation kinetics of ϵ COP-YFP and Arf1-CFP from the experiment in **b**. **d**, **e**, *IdIF* cells expressing ϵ COP-GFP were selectively photobleached within the

yellow outline after no treatment (**d**) or after treatment with AIF ($50 \mu\text{M AlCl}_3$, 0.1 mM NaF) for 10 min (**e**). Recovery into the bleached area was monitored by imaging cells with low illumination levels. **f**, **g**, CHO cells expressing only Arf1-GFP were selectively photobleached inside the yellow line after no treatment (**f**) or after treatment with AIF for 10 min (**g**). Recovery into the bleached area was monitored by imaging cells with low illumination levels. Scale bar, $5 \mu\text{M}$.

COPI to Golgi (process 2), 0.0384 s^{-1} for GTP hydrolysis by Golgi-bound Arf1 (process 4) and 0.0231 s^{-1} for loss of COPI from Golgi (process 6). Thus, the average residence time for coatomer on Golgi membranes is 43 s, in contrast with 26 s for Arf1. These residence times were the same whether calculated from photobleaching or from BFA-release experiments and are consistent with our current understanding that BFA inhibits solely the Arf1 exchange factor (process 1)^{10,12}.

We considered a simpler model in which all Arf1 used coatomer as an effector (that is, process 3 was eliminated). This model failed to account for the relative abundance of Arf1 and COPI on membrane and in cytoplasm. In another model we postulated the simultaneous release of Arf1 and coatomer to the cytoplasm, to correspond to the widespread assumption that uncoating is a consequence of Arf1-GTP hydrolysis². This model could not account for the slower coatomer dynamics observed during FRAP or BFA protocols. Thus, the model shown in Fig. 4c, which simultaneously accounts for all data sets (Fig. 4d), is our current working hypothesis.

It is widely believed that coatomer binds to membranes and then polymerizes into a coat. The coat then shapes the adherent membrane into a coated bud that gives rise to a coated vesicle (Fig. 5a). The vesicle then uncoats and all accumulated coatomer is released at once^{1,2}. Because of the sequential steps occurring before vesicle uncoating in this model, coatomer release should exhibit non-exponential kinetics (for example, it should show a shoulder or lag) (Supplementary Information). However, the observed release kinetics of coatomer from membranes (Fig. 4d) was exponential, leading to a possible alternative role for coatomer (Fig. 5b) in

which continuous Arf1-driven binding and dissociation of coatomer serves to generate kinetically stable membrane domains that give rise to coatomer-containing transport intermediates. In this model, unlike the sequential-binding model, coatomer exchange would continue even when vesicle formation is blocked.

Vesicular transport is known to be inhibited at temperatures below 15°C and nonexistent at 4°C (ref. 25). To evaluate the two models, we used FRAP to measure the kinetics of $\epsilon\text{COP-GFP}$ exchange between cytoplasmic and membrane pools at different temperatures. Both $\epsilon\text{COP-GFP}$ and Arf1-GFP were found to bind and release continuously from Golgi membranes at all temperatures down to 4°C , with no abrupt change in the kinetics of these processes on reaching temperatures (that is, 15°C) at which vesicle budding is slowed or inhibited²⁵ (Fig. 5c–e). At all temperatures, the release of $\epsilon\text{COP-GFP}$ from Golgi membranes was dependent on Arf1-GTP hydrolysis (not shown). The data thus indicate that the cycling of coatomer on and off membranes can be uncoupled from vesicle formation, and that feedback from productive vesicle budding is not necessary for coatomer dissociation.

On the basis of these findings, we favour the alternative model of coatomer activity with continuous membrane binding and release. Not only can it account for our observation that coatomer exhibits temperature-insensitive exponential release kinetics but it is consistent with other observations regarding the function and activity of coatomer. In the continuous binding and release model, coatomer, after being recruited by Arf1, could associate selectively with either protein or lipid species destined for transport^{23,26}, and coatomer–cargo complexes could cluster either because of direct inter-coatomer interactions or because of those induced via the mem-

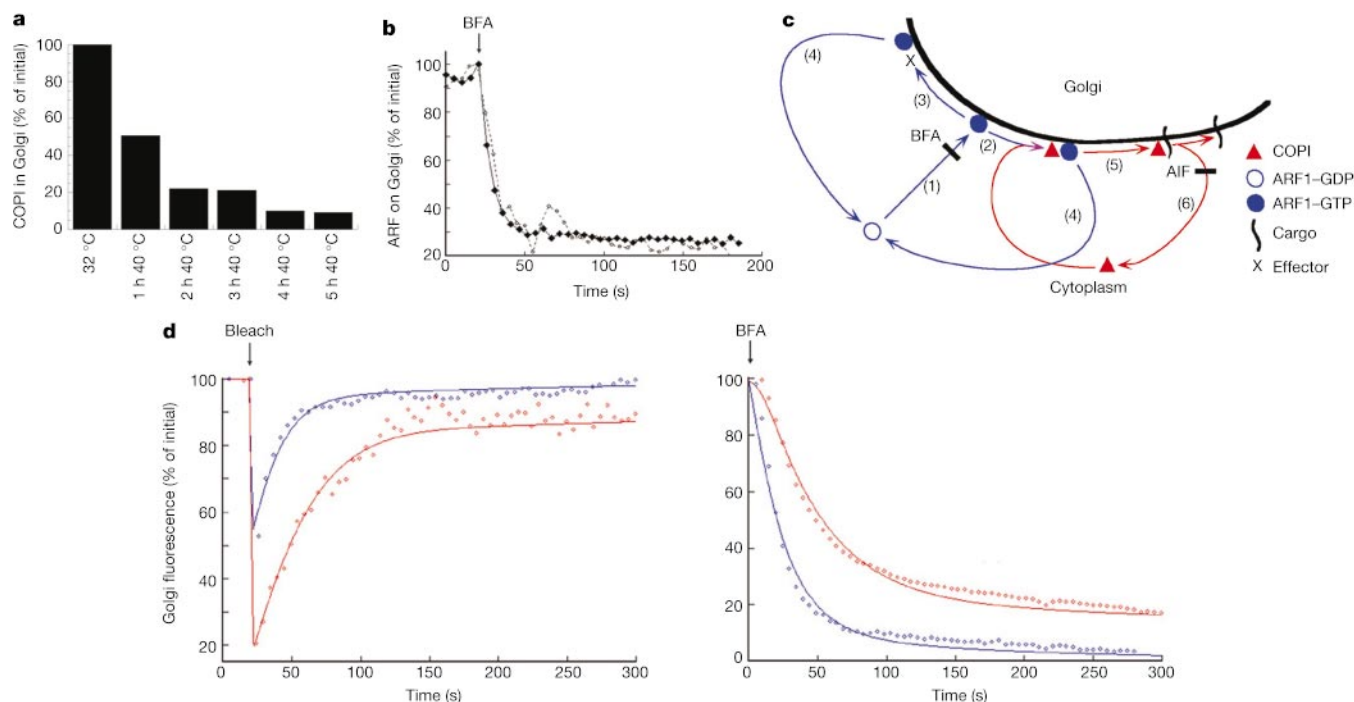


Figure 4 Kinetic modelling of coatomer and Arf1 activity and the effects of coatomer depletion on the rate of Arf1 dissociation from membranes. **a**, Conditions of coatomer depletion were determined by incubating, for increasing durations at 40°C , *IdIF* cells that were not transfected with any protein. The amount of Golgi-associated COPI (visualized by antibody labelling with a rabbit anti- βCOP and a rhodamine anti-rabbit secondary antibody) was determined. Each time point is an average of Golgi-localized fluorescence per cell from 15 fields from 3 different coverslips. **b**, *IdIF* cells were transfected with Arf1-GFP and incubated at 32°C (filled diamonds) or at 40°C for 3 h (open circles) to deplete COPI from Golgi membranes. BFA was then added to inhibit the binding of Arf1 to

membranes, and the release rate of Arf1-GFP fluorescence from Golgi was measured over time. **c**, Model of Arf1 and COPI membrane activity and turnover. **d**, Data sets from quantitative FRAP experiments measuring $\epsilon\text{COP-GFP}$ (red) and Arf1-GFP (blue) exchange on Golgi (see Supplementary Information) and the loss of $\epsilon\text{COP-YFP}$ (red) and Arf1-CFP (blue) from Golgi after the addition of BFA were simultaneously fitted to a mathematical formulation of the model shown in **c**. The smooth curves (red, COPI; blue, Arf1) represent the model solution obtained with parameter values (see the text) for the six processes in the model.

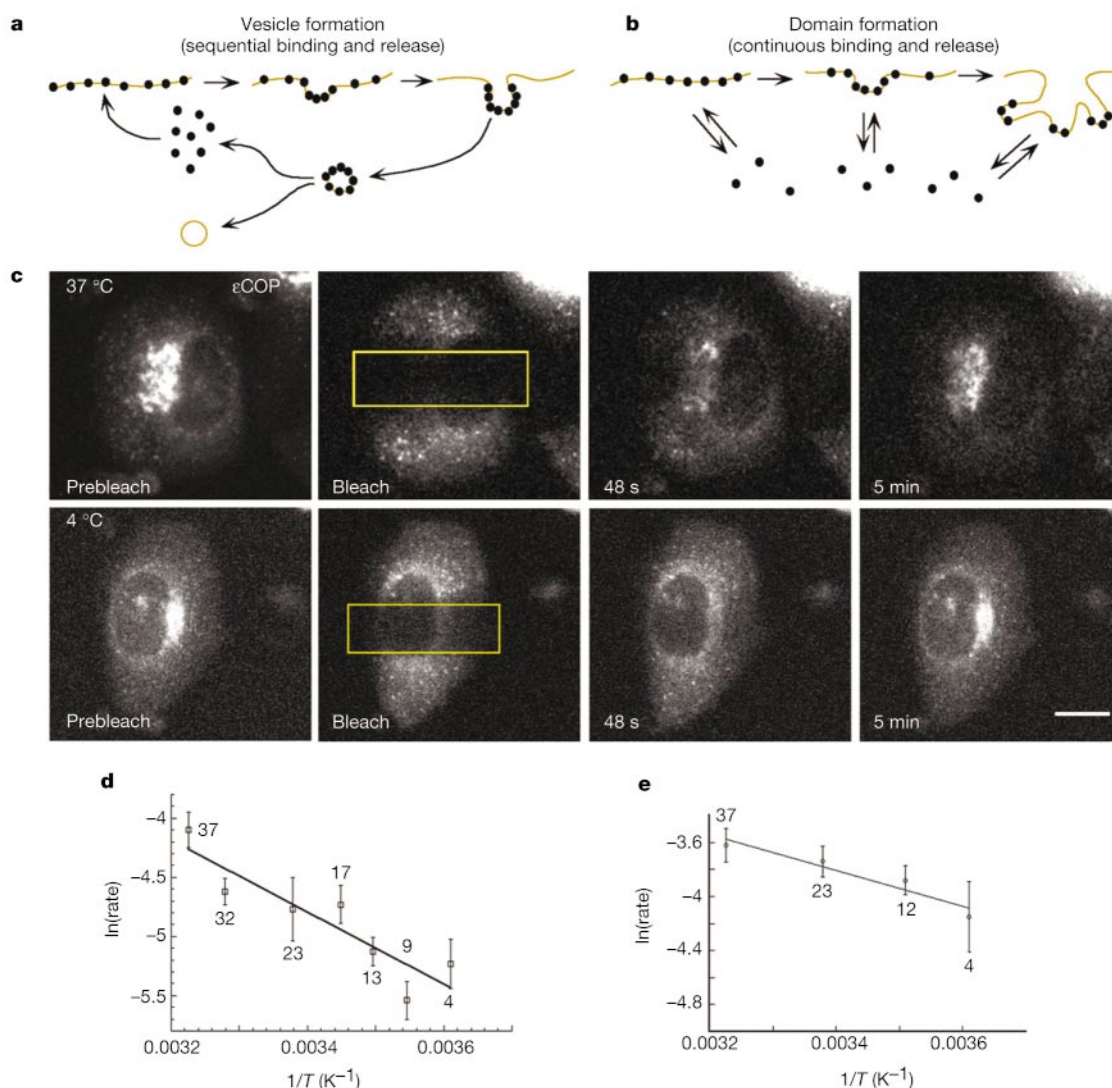


Figure 5 Testing different models of coatamer function on membranes. **a, b**, Two models for COPI function on Golgi membranes. In **a**, membrane binding and polymerization of COPI drives vesicle formation, with COPI dissociation from membranes occurring only after vesicle formation; in **b**, COPI binding and polymerization result in the formation of organized membrane domains that require continued binding and release of COPI to be maintained. **c**, FRAP of *IdIF* cell stably expressing ϵ COP–GFP on a microscope stage maintained at 37 °C or at 4 °C. Recovery of Golgi fluorescence in the bleached area (yellow box) occurred at both temperatures, indicating that membrane binding and release of

COPI occurs even when vesicle budding and fusion are inhibited. Scale bar, 5 μ M. (See Quicktime movies in Supplementary Information.) **d, e**, Arrhenius plots showing the effect of temperature on Golgi FRAP of ϵ COP–GFP in *IdIF* cells stably expressing ϵ COP–GFP (**d**) or in CHO cells transiently expressing Arf1–GFP (**e**). Numbers against points indicate temperatures in °C. The kinetics of recovery was fitted to a single-exponential curve and the mean rate coefficient was determined. Six to eight cells at each temperature were measured. The natural logarithm of the rate constant was then plotted against the reciprocal of the temperature in kelvins. Error bars indicate s.e.m.

branes. Continuing membrane binding and release of coatamer would allow the membrane at these sites to recruit cargo, alter their phospholipid composition and become larger, phase-separated domains (Fig. 5b). Accompanying these processes would be changes in membrane curvature and budding, which ultimately would transform the domains into either small coated vesicles² or larger transport intermediates^{4,27}. Throughout the process of membrane transformation, coatamer would continually be recruited to membranes and then released into the cytosol. Because clusters of coatamer complexes in the differentiated membrane domains would be maintained by the Arf1–coatamer binding and release cycle, the domains could last longer than the 43 s residence time of an individual coatamer complex on membranes. Detachment of a transport intermediate from the coatamer-coated domain would therefore occur on a completely different timescale from that of the release of individual coatamer complexes, which correlates with

topology changes underlying transport (that is, tubulation, fission and vesiculation) rather than uncoating.

Our observation that the activity of coatamer is not directly coupled to the formation of coated vesicles is consistent with recent FRAP data on the behaviour of clathrin, which has revealed large-scale exchange of clathrin even when clathrin-mediated endocytosis is inhibited²⁸. Thus, coat proteins in general might undergo continuous exchange on and off membranes. As with coatamer, further investigation will help to determine whether this process is necessary to generate and maintain transport competent membrane domains, which could also function in membrane sorting and signalling. □

Methods

Cloning and expression of complementary DNAs

Arf1[Q71L], ϵ COP–GFP, ϵ COP–YFP, p58–CFP and VSVG–CFP plasmids were described

previously^{13,29}. Arf1-CFP and Arf1-YFP were derived by the exchange of GFP for its spectral variant from Arf1-GFP¹¹. DNAs were transiently transfected into cells by using FuGene 6 Transfection Reagent (Roche Molecular Biochemicals). Cell lines stably expressing eCOP-GFP and YFP were created by transient transfection followed by selection with G418, or in stably expressing *ldlf* cells selection was solely by ability to grow at 40 °C. Except where indicated, all experiments were performed in stable *ldlf* cell lines expressing eCOP-GFP/YFP at 40 °C, in *ldlf* cells transiently transfected with Arf1-GFP/CFP, and in wild-type Chinese hamster ovary (CHO) cells.

Fluorescence microscope systems, imaging and photobleaching analysis

Cells on coverslips or in LabTek chambers (Nalge Nunc) were imaged in buffered medium with a Zeiss 510 confocal microscope with a stage heated to 37 °C and fitted with an additional krypton laser with a 413-nm line for CFP/YFP double labelling or on a similarly configured Zeiss 410 confocal system. The time-lapse sequence shown in Fig. 1d and Fig. 1e was taken with a high-frame-rate spinning-disc confocal system, the Ultraview (lent by EG&G Wallac). Quantitative Golgi FRAP of eCOP-GFP was performed as described in Supplementary Information.

Received 2 November 2001; accepted 26 January 2002.

- Schekman, R. & Orci, L. Coat proteins and vesicle budding. *Science* **271**, 1526–1533 (1996).
- Rothman, J. E. & Wieland, F. T. Protein sorting by transport vesicles. *Science* **272**, 227–234 (1996).
- Guo, Q., Penman, M., Trigatti, B. L. & Krieger, M. A single point mutation in ϵ -COP results in temperature-sensitive, lethal defects in membrane transport in a Chinese hamster ovary cell mutant. *J. Biol. Chem.* **271**, 11191–11196 (1996).
- Presley, J. F. *et al.* ER-to-Golgi transport visualized in living cells. *Nature* **389**, 81–85 (1997).
- Lahtinen, U., Dahllof, B. & Saraste, J. Characterization of a 58 kDa *cis*-Golgi protein in pancreatic exocrine cells. *J. Cell Sci.* **103**, 321–333 (1992).
- Shima, D. T., Scales, S. J., Kreis, T. E. & Pepperkok, R. Segregation of COPI-rich and anterograde-cargo-rich domains in endoplasmic-reticulum-to-Golgi transport complexes. *Curr. Biol.* **9**, 821–824 (1999).
- Klumpperman, J. *et al.* The recycling pathway of protein ERGIC-53 and dynamics of the ER-Golgi intermediate compartment. *J. Cell Sci.* **111**, 3411–3425 (1998).
- Lippincott-Schwartz, J., Snapp, E. & Kenworthy, A. Studying protein dynamics in living cells. *Nature Rev. Mol. Cell Biol.* **2**, 444–456 (2001).
- Dascher, C. & Balch, W. E. Dominant inhibitory mutants of ARF1 block endoplasmic reticulum to Golgi transport and trigger disassembly of the Golgi apparatus. *J. Biol. Chem.* **269**, 1437–1448 (1994).
- Klausner, R. D., Donaldson, J. G. & Lippincott-Schwartz, J. Brefeldin A: Insights into the control of membrane traffic and organelle structure. *J. Cell Biol.* **116**, 1071–1080 (1992).
- Vasudevan, C. *et al.* The distribution and translocation of the G protein ADP-ribosylation factor 1 in live cells is determined by its GTPase activity. *J. Cell Sci.* **111**, 1277–1285 (1998).
- Peyroche, A. *et al.* Brefeldin A acts to stabilize an abortive ARF-GDP-Sec7 domain protein complex: Involvement of specific residues of the Sec7 domain. *Mol. Cell* **3**, 275–285 (1999).
- Teal, S. B., Hsu, V. W., Peters, P. J., Klausner, R. D. & Donaldson, J. G. An activating mutation in ARF1 stabilizes coatamer binding to Golgi membranes. *J. Biol. Chem.* **269**, 3135–3138 (1994).
- Goldberg, J. Decoding of sorting signals by coatamer through a GTPase switch in the COPI coat complex. *Cell* **100**, 671–679 (2000).
- Szafer, E., Rotman, M. & Cassel, D. Regulation of GTP hydrolysis on ADP-ribosylation factor-1 at the Golgi membrane. *J. Biol. Chem.* **276**, 47834–47839 (2001).
- Donaldson, J. G., Kahn, R. A., Lippincott-Schwartz, J. & Klausner, R. D. Binding of ARF and β -COP to Golgi membranes: possible regulation by a trimeric G protein. *Science* **254**, 1197–1199 (1991).
- Kahn, R. A. Fluoride is not an activator of the smaller (20–25 kDa) GTP-binding proteins. *J. Biol. Chem.* **266**, 15595–15597 (1991).
- Chardin, P. *et al.* A human exchange factor for ARF contains Sec7- and pleckstrin-homology domains. *Nature* **384**, 481–484 (1996).
- Peyroche, A., Paris, S. & Jackson, C. L. Nucleotide exchange on ARF mediated by yeast Gea1 protein. *Nature* **384**, 479–481 (1996).
- Zhao, L. *et al.* Direct and GTP-dependent interaction of ADP ribosylation factor 1 with coatamer subunit β . *Proc. Natl Acad. Sci. USA* **94**, 4418–4423 (1997).
- Puertollano, R., Randazzo, P. A., Presley, J. F., Hartnell, L. M. & Bonifacio, J. S. The GGAs promote ARF-dependent recruitment of clathrin to the TGN. *Cell* **105**, 93–102 (2001).
- Makler, V., Cukierman, E., Rotman, M., Admon, A. & Cassel, D. ADP-ribosylation factor-directed GTPase-activating protein. Purification and partial characterization. *J. Biol. Chem.* **270**, 5232–5237 (1995).
- Sohn, K. *et al.* A major transmembrane protein of Golgi-derived COPI-coated vesicles involved in coatamer binding. *J. Cell Biol.* **135**, 1239–1248 (1996).
- Lanoix, J. *et al.* GTP hydrolysis by arf-1 mediates sorting and concentration of Golgi resident enzymes into functional COP I vesicles. *EMBO J.* **18**, 4935–4948 (1999).
- Bremser, M. *et al.* Coupling of coat assembly and vesicle budding to packaging of putative cargo receptors. *Cell* **96**, 495–506 (1999).
- Cosson, P. & Letourneur, F. Coatamer interaction with di-lysine endoplasmic reticulum retention motifs. *Science* **263**, 1629–1631 (1994).
- Lippincott-Schwartz, J., Roberts, T. H. & Hirschberg, K. Secretory protein trafficking and organelle dynamics in living cells. *Annu. Rev. Cell Dev. Biol.* **16**, 557–589 (2000).
- Wu, X. *et al.* Clathrin exchange during clathrin-mediated endocytosis. *J. Cell Biol.* **155**, 291–300 (2001).
- Ward, T. H., Polishchuk, R. S., Caplan, S., Hirschberg, K. & Lippincott-Schwartz, J. Maintenance of Golgi structure and function depends on the integrity of ER export. *J. Cell Biol.* **155**, 557–570 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank D. Hailey for help in the analysis of Arf1-GFP in yeast cells; M. Krieger for his gift of *ldlf* cells; F. Wieland and J. E. Rothman for eCOP cDNA; J. Donaldson for Arf1 [Q71L] cDNA; G. Romero for the Arf1-GFP plasmid; EG&G Wallac for the use of their Ultravision spinning disc confocal system; and J. Donaldson, J. Bonifacio, C. Jackson, K. Hirschberg, B. Nichols, A. Kenworthy, N. Cole and H. Radhakrishna for valuable discussion. J.F.P. was supported in part by a grant from the Canadian Institutes of Health Research.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J. L. S. (e-mail: jlippin@helix.nih.gov).

Microtubule basis for left-handed helical growth in *Arabidopsis*

Siripong Thitamadee, Kazuko Tuchihiro & Takashi Hashimoto

Graduate School of Biological Sciences, Nara Institute of Science and Technology, Takayama 8916-5, Ikoma 630-0101, Japan

Left-right asymmetry in plants can be found in helices of stalks, stems and tendrils, and in fan-like petal arrangements. The handedness in these asymmetric structures is often fixed in given species, indicating that genetic factors control asymmetric development¹. Here we show that dominant negative mutations at the tubulin intradimer interface of α -tubulins 4 and 6 cause left-handed helical growth and clockwise twisting in elongating organs of *Arabidopsis thaliana*. We demonstrate that the mutant tubulins incorporate into microtubule polymers, producing right-handed obliquely oriented cortical arrays, in the root epidermal cells. The cortical microtubules in the mutants had increased sensitivity to microtubule-specific drugs. These results suggest that reduced microtubule stability can produce left-handed helical growth in plants.

Most twining plants, as they grow upward around sticks, trees, or other plants, coil in right-handed helices. Others coil the opposite way, and some exceptional species have genetically determined left- and right-handed varieties¹. In contrast, the handedness of corolla contortion is either random in a species or fixed in a species or larger taxon. The fixed pattern mostly occurs in (but is not restricted to) the asterids, in which both clockwise and anticlockwise petal arrangements are seen in taxonomically loosely related groups². Widespread distribution of helical structures with fixed handedness in diverse taxa implies that relatively small numbers of genes are responsible for the generation of such asymmetry.

Wild-type *Arabidopsis thaliana* axial organs do not twist during normal elongation growth and flowers are radially symmetrical (Fig. 1a, f, k). This symmetry can be broken, however, by mutations, including the right-handed helical growth mutants, *spiral1* (*spr1*) and *spr2* (ref. 3). In a search for suppressor mutants of the right-hand twisting mutant *spr1*, we identified two independent mutant loci, *lefty1* and *lefty2*. When outcrossed to remove the *spr1* mutation, both mutants display prominent left-handed helical growth phenotypes. The twisting phenotypes are somewhat stronger in *lefty2* than in *lefty1*. Petioles and petals of *lefty* mutants twist to give an appearance of clockwise bending when rosette leaves and flowers are viewed from above (Fig. 1g, h, l, m). The epidermal cell files of *lefty* hypocotyls and roots (Fig. 1q, r, t, u) form left-handed (S-form) helices, in contrast to straight wild-type cell files (Fig. 1p, s). The epidermal cell files of *lefty* roots begin to skew at the region where

the first root hair is emerging, and the length of the region between the proximal end of the root cap and the first root hair initial, an indicator of the elongation zone⁴, is shorter in *lefty* roots than in wild-type roots (Fig. 1v). When *lefty* seedlings are cultured vertically against an impenetrable agar surface, this left-handed twisting of epidermal files generates a rightward bend, and primary roots skew sharply toward the left side of the plate, as viewed from above (Fig. 1b, c). Despite their twisting habit, *lefty* plants lack other vegetative or reproductive abnormalities.

We determined that the *lefty1* and *lefty2* mutations result from single nucleotide exchanges in α -tubulin genes *TUA6* and *TUA4*, respectively. These two genes have close sequence identity and both mutations encode an exchange of the serine at amino-acid position 180 with a phenylalanine (Fig. 2). This residue is conserved in fungal, animal, and plant α -tubulins as either a serine or an alanine⁵, and is predicted to be located in the T5 loop at the

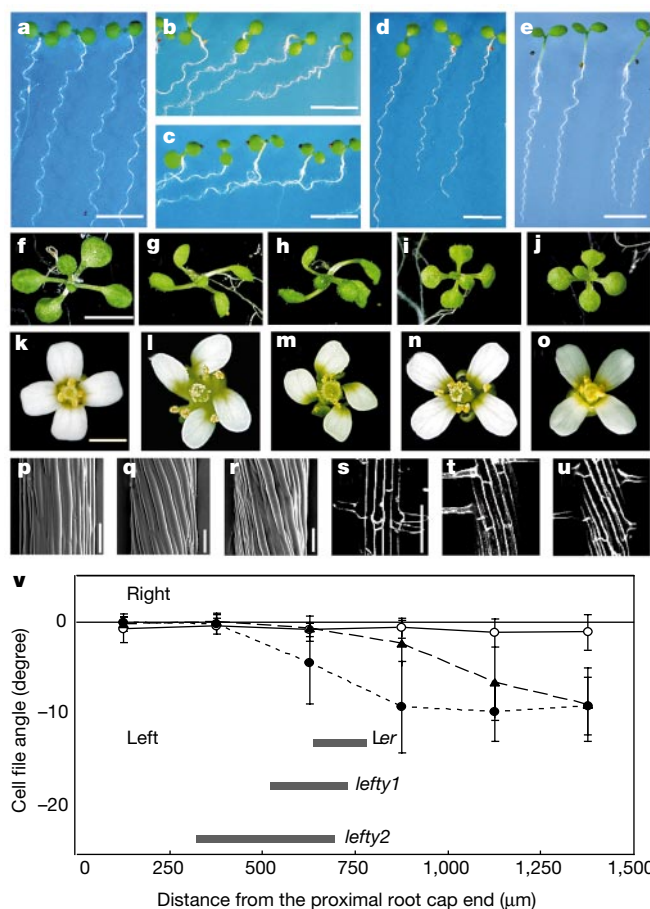


Figure 1 Helical growth phenotypes of *lefty* mutants and their suppressors. **a–e**, Seven-day-old seedlings grown on vertical hard-agar surfaces as viewed from above the agar surface. **f–j**, Two-week-old seedlings with cotyledons and the first 2–4 true leaves. Scale bars for **a–j**, 5 mm. **k–o**, Flowers. Scale bars for **k–o**, 1 mm. **p–r**, Scanning electron micrographs of 7-day-old light-grown hypocotyls. **s–u**, Montage of propidium iodide stained confocal sections showing differentiated regions of 5-day-old primary roots. Scale bars for **p–u**, 100 μm. *Ler* wild type (**a, f, k, p, s**), *lefty1* (**b, g, l, q, t**), *lefty2* (**c, h, m, r, u**), *lefty1R1* (**d, i, n**), and *lefty2R1* (**e, j, o**). **v**, Skewed angles of root epidermal cell files deviating from the main axis of primary roots, with right- and left-handed helices being designated as positive and negative, respectively. Cells at the indicated position from the proximal end of the root cap were examined. Open circles with solid lines, *Ler*; filled triangles with dashed lines, *lefty1*; filled circles with dotted lines, *lefty2*. The regions of the first visible root hair formation were shown in horizontal bars for wild type (832 ± 66 ; mean $\pm$ s.d.), *lefty1* (741 ± 101), and *lefty2* (624 ± 176).

Table 1 Drug sensitivity of wild-type and *lefty* seedlings

	ID ₅₀ (mg l ⁻¹)				
	Wild type	<i>LEFTY1/lefty1</i>	<i>lefty1/lefty1</i>	<i>LEFTY2/lefty2</i>	<i>lefty2/lefty2</i>
Taxol	0.5	0.19	0.13	0.23	0.17
Propyzamide	1.32	0.98	0.60	0.88	0.54

Sensitivity data is shown for wild-type, *lefty* homozygotes and their heterozygotes to microtubule-interacting drugs. The ID₅₀ value was calculated from the dose-response curves of seedlings. Hypocotyl length of 5-day-old etiolated seedlings was the measure for Taxol treatment, whereas primary root length of 7-day-old light-grown seedlings was scored for propyzamide treatment.

longitudinal contact region with β -tubulin within tubulin dimers⁶. Substituting a bulkier amino acid at this position is expected to affect the structure and/or stability of tubulin dimers and hence polymer function.

The orientation of cortical microtubules is generally believed to specify the orientation of nascent cellulose microfibrils, thereby dictating the direction of cell elongation⁷. We examined cortical microtubules underneath the outer tangential wall of epidermal cells in the root elongation zone (Fig. 3). Wild-type cortical microtubule arrays were aligned almost transverse to the long axis of the cell (Fig. 3b), whereas the arrays in *lefty1* and *lefty2* formed right-handed helices (Fig. 3c). The right-handed helical arrays were observed in the *lefty* epidermal cells well before the visible left-handed slanting of cell files began at the end of the elongation zone (Fig. 3a, also see Fig. 1v), indicating that the oblique array arrangement is not the direct consequence of twisted epidermal cells. When seedling roots were treated with 0.15 μM oryzalin, a microtubule-depolymerizing drug, wild-type root epidermal cells retained cortical microtubules but these microtubules were oriented in right-handed helices (Fig. 3d). At the same concentration of oryzalin, *lefty* cortical microtubules were depolymerized or heavily fragmented (Fig. 3e), indicating that their stability was low compared with wild-type microtubules.

We tested the responses of *lefty* mutant seedlings to the tubulin-specific drugs⁸ Taxol, oryzalin, propyzamide, and RH-4032 (a potent propyzamide analogue⁹) (Table 1, and Supplementary Information). These drugs inhibited root and hypocotyl elongation and caused isotropic expansion of epidermal cells at lower doses in *lefty* mutants than in wild-type seedlings. The responsiveness of heterozygotes to tubulin-specific drugs was intermediate between those of *lefty* homozygotes and the wild type. This semi-dominance was confirmed in a root slanting assay (Supplementary Information). The *lefty* heterozygotes and the Landsberg *erecta* (*Ler*) parent ecotype are indistinguishable morphologically without drug treatment. In the *spr1* mutant background, however, heterozygous *lefty1* and *lefty2* mutants partially suppress the right-handed twisting of *spr1*.

We next chemically mutagenized *lefty1* and *lefty2*, and screened for suppressor mutants. One *lefty1* suppressor (*lefty1R1*) and three

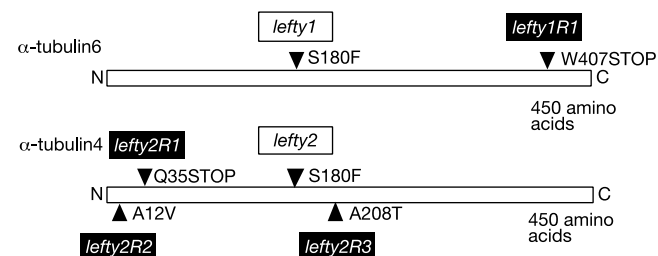


Figure 2 Mutation points in *lefty* and their suppressors in amino-acid sequences of α -tubulin6 and α -tubulin4. We note that each suppressor has the original S180F mutation in addition to the indicated second-site mutation.

lefty2 suppressors (*lefty2R1* to *lefty2R3*) are reported here. These suppressor mutants were indistinguishable from the parent ecotype in their growth phenotypes (Fig. 1d, e, i, j, n, o), and in their responses to microtubule-specific drugs (data not shown). All suppressor mutations were found in the original *LEFTY* loci (Fig. 2). The *lefty1* suppressor mutant *lefty1R1* carries a second site mutation, introducing a stop codon at position 407, and

resulting in a truncated TUA6 protein. The missing carboxy terminus includes part of the loop connecting helix 11 and helix 12 that is crucial for the interaction between α - and β -tubulin monomers⁶. Second site mutations were also found in the three *lefty2* suppressor mutants, with a stop codon introduced at position 35 in *lefty2R1*, and amino-acid substitutions at position 12 and 208 in *lefty2R2* and *lefty2R3*, respectively. This suppressor analysis suggests that although *lefty1* and *lefty2* are dominant negative mutations of α -tubulins, their null alleles are phenotypically indistinguishable from the wild type.

Suppression of the *lefty* phenotypes by their null alleles suggests that incorporation of the mutant tubulin into microtubule polymers alters the microtubule properties. To assess this possibility, we tagged the wild-type and *lefty1* versions of TUA6 at their C termini with the *c-myc* epitope, and expressed these constructs under the 35S promoter of cauliflower mosaic virus in transgenic *Arabidopsis* plants. Transgenic lines expressing wild-type TUA6-myc had wild-type phenotypes, whereas seedlings transformed with the mutant TUA6^{S180F}-myc reproduced all *lefty* phenotypes, including increased drug sensitivity (data not shown). Double immunocytochemistry with antibodies raised against α -tubulin and *c-myc* epitope indicated that both myc-tagged proteins were uniformly incorporated into cortical microtubules in root epidermal cells (Fig. 4). However, the arrays containing TUA6^{S180F}-myc formed right-handed helices, in contrast to the transversely aligned arrays in

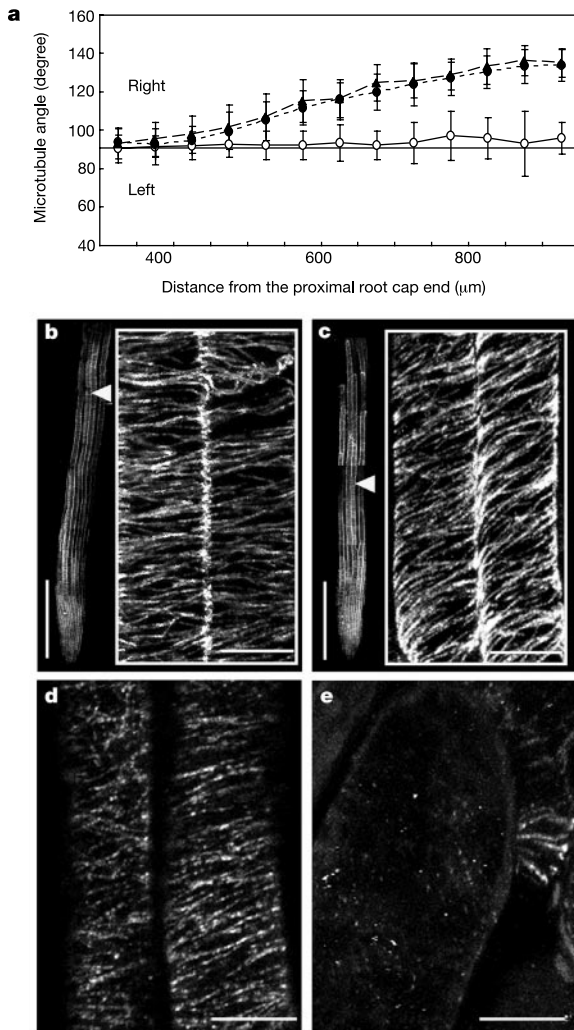


Figure 3 The *lefty* mutant cortical microtubules are arranged in right-handed helices and are more sensitive to microtubule-destabilizing drugs. **a**, Anti-tubulin immunofluorescence of root epidermal cells enabled measurement of angular deviation of microtubules from the transverse axis throughout the root elongation zone. Angles above 90° indicate right-handed helices. Microtubule orientation in *Ler* cells (open circles with solid lines) deviated relatively little from the transverse axis, while microtubules in the *lefty1* (filled triangles with dashed lines) and *lefty2* (filled circles with dotted lines) mutants formed right-handed helical arrays throughout the elongation zone. **b**, Anti-tubulin immunofluorescence in wild type (*Ler*) root tip demonstrates transverse cortical microtubules in a cell in the distal elongation zone. **c**, In *lefty1*, cortical microtubules are right-handed oblique even in the mid-elongation zone. Arrowheads indicate position of the cell depicted in the inset relative to the root cap. Scale bars, 200 μm for low-magnification images in **b** and **c**, 10 μm for insets. **d**, **e**, Root epidermal cells in the late elongation zone from 5-day-old wild-type (**d**) and *lefty2* (**e**) seedlings grown in the presence of 0.15 μM oryzalin. The wild-type microtubules are oriented in slightly right-handed oblique array, whereas the *lefty2* microtubules are heavily disrupted with only a few intact microtubules detected. Scale bars in **d** and **e**, 10 μm. Roots were fixed and stained as reported before^{3,11}.

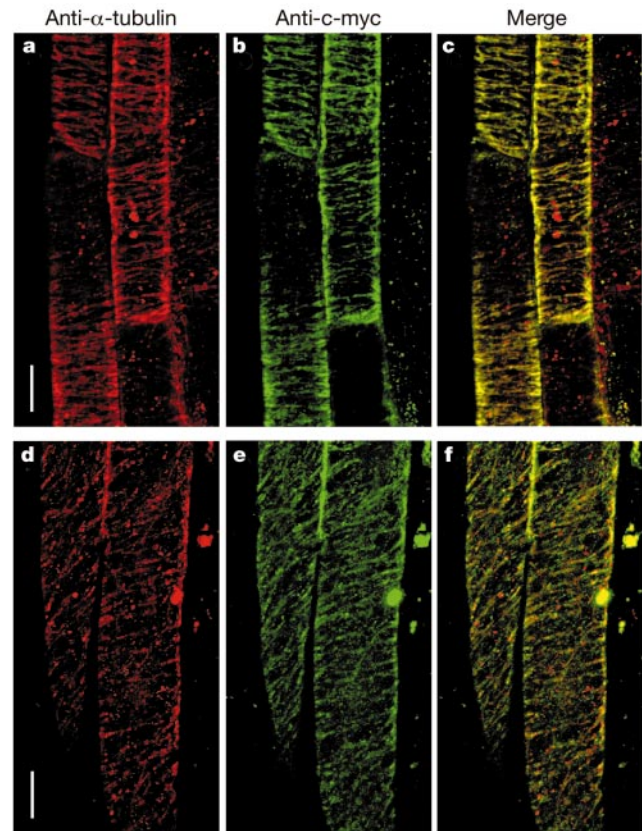


Figure 4 Incorporation of *lefty*-type tubulin into microtubule arrays. Root epidermal cells at the elongation zone were stained. **a–c**, Transgenic plant expressing TUA6-myc. **d–f**, Transgenic plant expressing *lefty*-type TUA6^{S180F}-myc. **a**, **d**, Anti- α -tubulin primary antibody and Alexa Fluor 568-coupled secondary antibody were used to visualize cortical microtubules. **b**, **e**, Anti-*c-myc* primary antibody and FITC-conjugated secondary antibody were used to show the distribution of the Myc-tagged proteins. **c**, **f**, Merged images of the left two panels. Colocalized signals appear in yellow. Scale bars, 10 μm.

the TUA6-myc control. Mitotic spindles and phragmoplasts were also stained with myc-specific antibody, indicating incorporation of the tagged tubulins into these arrays, but these microtubule structures were indistinguishable in appearance in both transgenic lines (data not shown). Cell division patterns were normal in both lines. Thus, the function and organization of cortical microtubule arrays may be more sensitive to incorporation of the aberrant *lefty* tubulin than mitotic and cytokinetic microtubule structures.

The *Arabidopsis* genome contains six expressed α -tubulin genes. Among them, *TUA2* and *TUA4* encode identical proteins, and *TUA6* differs from *TUA2/4* at only two amino-acid residues¹⁰. These three tubulin genes are expressed in overlapping patterns in roots, leaves and flowers (ref. 10 and our unpublished results). Null *TUA4* and *TUA6* mutations, identified as *lefty* suppressors, had no obvious phenotypes, probably because other α -tubulin isoforms compensated for their functions. The tubulins encoded by *TUA4* and *TUA6* have similar functions, as indicated by the observation that the gain-of-function alleles of *TUA4* and *TUA6* have similar phenotypes, as do the loss-of-function alleles of these two genes. The presence of aberrant *lefty*-type tubulin in microtubule structures (Fig. 4) suggests that *lefty* α -tubulins readily form dimers with β -tubulin, and that these mutant tubulin dimers co-assemble with wild-type tubulin dimers into microtubule protofilaments.

The right-handed oblique orientation of *lefty* cortical microtubules could be the principal cause of the organ twisting, because they are approximately perpendicular to the left-handed oblique long axis of the epidermal cell files. In this context, determining the orientation of cellulose microfibrils in helically growing *lefty* cells is an important goal, though microfibril alignment can, at least in some cases, be independent of microtubules¹¹. Just why microtubule polymers containing mutant α -tubulins should form right-handed arrays is still unclear, though microtubules in *Arabidopsis* root epidermal cells consistently form a transient right-handed helix during the final stage of elongation¹² at a time when they may be relatively unstable. What is clear, however, is that reducing cortical microtubule stability results in left-handed organ twisting in *Arabidopsis*. This is demonstrated either with low doses of microtubule-targeted drugs³ or in the temperature-sensitive *mor1* mutant¹³. The *MOR1* gene encodes a high-relative-molecular-mass microtubule-associated protein that is essential for cortical microtubule organization. Several other *Arabidopsis* helical growth mutants with fixed handedness have also been reported^{3,14,15}, but the molecular and cellular bases of their asymmetrical growth are unknown. Our results suggest that skewed arrangement of cortical microtubules offers a potential mechanism to generate the left-right asymmetry in elongating axial organs, and provides a possible molecular explanation not only for chirality establishment in *Arabidopsis* twisting mutants but also for the natural asymmetry seen in tendrils, vines and other twisting plant organs. □

Methods

Mutant screening

We mutagenized *spr1-1* in the *Ler* background with ethyl methane sulphonate (EMS), and screened the M2 populations using a root slanting assay on vertically positioned hard-agar plates³. *lefty1* and *lefty2* were easily identified because their roots skewed sharply to the left, as opposed to the right in *spr1* roots, on the agar surface. The original isolates were backcrossed to *Ler* wild type, and *lefty* mutants without the *spr1* mutation were obtained. A 0.6-kb deletion in the *SPR1* gene was used to detect the *spr1-1* mutation by genomic polymerase chain reaction PCR (our unpublished results). Suppressors of *lefty1* were similarly isolated after screening *lefty1spr1* M2 lines treated with EMS for seedlings with right-handed slanting. After the original suppressors were backcrossed to *lefty1*, *lefty1* suppressors without the *spr1* mutation were obtained. Further data on the genetics of *lefty* and *spr1* can be obtained in the Supplementary Information. For *lefty2* suppressor screening, EMS-treated *lefty2* M2 lines were grown on a plate containing propyzamide analogue RH-4032 (ref. 8) at 10 nM and screened for seedlings with wild-type sensitivity to the drug.

Molecular cloning of *LEFTY* loci

To generate lines for ecotype-specific marker mapping of mutant loci, *lefty1spr1-1* and

lefty2spr1-1 were crossed to *spr1-2* (Columbia background). Because *lefty* mutations were epistatic to *spr1* and semi-dominant in the *spr1* background (Supplementary Information), *spr1* segregants that were wild type at *LEFTY* loci were identified in F₂ seedlings by screening for right-handed root slanting on a hard-agar surface. DNA extracted from wild-type homozygotes at *LEFTY* loci was used to map *LEFTY* using cleaved amplified polymorphic sequence markers¹⁶ and simple sequence length polymorphism markers¹⁷. We examined 3,310 chromosomes and 800 chromosomes, and found that *LEFTY1* mapped to the 110-kb interval at ~49 cM on chromosome 4, within the bacterial artificial chromosome (BAC) ATFCOA, and that *LEFTY2* mapped to the 400-kb interval at ~7 cM on chromosome 1 between BACs F19P19 and yUP8H12 on the top of chromosome 1. Genomic PCR product sequence analysis of *TUA6* and *TUA4* genes located in these regions identified point mutations in the *lefty* mutants as well as in their suppressors.

Transgenic *Arabidopsis* lines

An *NdeI*–*HindIII* genomic fragment containing the entire coding region and the single intron of *AtTUA6* was excised from the g20182 cosmid clone, and subcloned into pBluescript II. A *SmaI* site was introduced at position –15 before the first ATG start codon, and an *XbaI* site was introduced immediately before the TAA stop codon. An oligonucleotide encoding *c-myc* epitope and a stop codon was then introduced between the *XbaI* and *SacI* sites. Separately, genomic PCR was used to amplify the *TUA6* region from *lefty1* to replace the wild-type gene as an *AccI* fragment. The resultant *TUA6-myc* constructs were cloned as *SmaI*–*SacI* fragments into pBI121. *A. thaliana* plants (Columbia ecotype) were transformed by floral dipping into *Agrobacterium tumefaciens* GV3101 strains containing the binary vectors. Among dozens of transformants with similar phenotypes, two T3 lines homozygous for a single transferred DNA insert were selected for each construct and these were used for microtubule and morphological phenotype analysis.

Received 17 December 2001; accepted 18 February 2002.

- Gardner, M. in *The New Ambidextrous Universe: Symmetry and Asymmetry from Mirror Reflections to Superstrings* 54–57 (Freeman, New York, 1990).
- Endress, P. K. Symmetry in flowers: diversity and evolution. *Int. J. Plant Sci.* **160**, S3–S23 (1999).
- Furutani, I. *et al.* The *SPIRAL* genes are required for directional control of cell elongation in *Arabidopsis thaliana*. *Development* **127**, 4443–4453 (2000).
- Dolan, L. *et al.* Cellular organisation of the *Arabidopsis thaliana* root. *Development* **119**, 71–84 (1993).
- Burns, R. G. & Surridge, C. D. in *Microtubules* 3–31 (Wiley-Liss, New York, 1994).
- Nogales, E., Wolf, S. G. & Downing, K. H. Structure of the $\alpha\beta$ tubulin dimer by electron crystallography. *Nature* **391**, 199–203 (1998).
- Griddings, T. H. Jr & Stachelin, L. A. in *The Cytoskeletal Basis of Plant Growth and Form* 85–100 (Academic, San Diego, 1991).
- Morejohn, L. C. in *The Cytoskeletal Basis of Plant Growth and Form* 29–43 (Academic, San Diego, 1991).
- Young, D. H. & Lewandowski, V. T. Covalent binding of the benzamide RH-4032 to tubulin in suspension-cultured tobacco cells and its application in a cell-based competitive-binding assay. *Plant Physiol.* **124**, 115–124 (2000).
- Kopczak, S. D., Haas, N. A., Hussey, P. J., Silflow, C. D. & Snustad, D. P. The small genome of *Arabidopsis* contains at least six expressed α -tubulin genes. *Plant Cell* **4**, 539–547 (1992).
- Baskin, T. I. On the alignment of cellulose microfibrils by cortical microtubules: a review and a model. *Protoplasma* **215**, 150–171 (2001).
- Sugimoto, K., Williamson, R. E. & Wasterneys, G. O. New techniques enable comparative analysis of microtubule orientation, wall texture, and growth rate in intact roots of *Arabidopsis*. *Plant Physiol.* **124**, 1493–1506 (2000).
- Whittington, A. T. *et al.* MOR1 is essential for organizing cortical microtubules in plants. *Nature* **411**, 610–613 (2001).
- Rutherford, R. & Masson, P. H. *Arabidopsis thaliana sku* mutant seedlings show exaggerated surface-dependent alteration in root growth vector. *Plant Physiol.* **111**, 987–998 (1996).
- Marinelli, B., Gomasasca, S. & Soave, C. A pleiotropic *Arabidopsis thaliana* mutant with inverted root chirality. *Planta* **202**, 196–205 (1997).
- Konieczny, A. & Ausubel, F. M. A procedure of mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* **4**, 403–410 (1993).
- Bell, C. J. & Ecker, J. R. Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19**, 137–144 (1994).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank V. Lewandowski for RH-4032, G. Wasterneys for valuable comments and suggestions on the manuscript, and ABRC for the g20182 clone. This study was in part supported by grants from the Ministry of Education, Science and Culture of Japan to T.H.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.H. (e-mail: hasimoto@bs.aist-nara.ac.jp).

Cnd2 has dual roles in mitotic condensation and interphase

Nobuki Aono, Takashi Sutani*, Takeshi Tomonaga*, Satoru Mochida & Mitsuhiro Yanagida

Graduate School of Biostudies, Department of Gene Mechanisms, Kyoto University, Kitashirakawa-Oiwakecho, Sakyo-ku, Kyoto 606-8582, Japan

Chromosome condensation requires condensin^{1–4}, which comprises five subunits^{5,6}. Two of these subunits—both being structural maintenance of chromosome (SMC) proteins—are coiled-coils with globular terminal domains that interact with ATP and DNA. The remaining three, non-SMC subunits also have essential, albeit undefined, roles in condensation. Here we report that Cnd2 (ref. 6), a non-SMC subunit of fission yeast similar to *Drosophila* Barren⁷ and the budding yeast protein Brn1 (refs 8, 9), is required for both interphase and mitotic condensation. In *cnd2-1* mutants, ultraviolet-induced DNA damage is not repaired, and cells arrested by hydroxyurea do not recover. A definitive defect of interphase is abolishment of Cds1 (a checkpoint kinase) activation in the presence of hydroxyurea in both *cnd2-1* mutant cells and in cells where other condensin subunits have been genetically disrupted. In the absence of hydroxyurea, a G2 checkpoint delay occurred in *cnd2-1* mutants in a manner dependent on Cds1 and ATM-like Rad3, but not Chk1 (refs 10–13), before the mitotic condensation defect. Furthermore, *cnd2-1* was synthetic-lethal with mutations of excision repair, RecQ helicase and DNA replication enzymes. These interphase and mitotic defects provide insight into the mechanistic role of non-SMC subunits that interact with the globular SMC domains in the heteropentameric holocomplex¹⁴.

Screening a collection of fission yeast, temperature-sensitive mutants¹⁵ cultured at 36 °C yielded ten strains defective in mitotic chromosome condensation. Plasmids pCND1, pCND2 or pCND3 (ref. 6) were introduced into these strains by transformation, and the resulting transformants were plated at 36 °C. One strain, designated *cnd2-1*, was rescued at 36 °C, but only by pCND2 (Supplementary Information Fig. 1a). Mutant cells stained by 4,6-diamidino-2-phenylindole (DAPI; see Supplementary Information Fig. 1b) produced a phenotype highly resembling the condensation defect of *cnd2* null or *cut14-208* mutants (refs 2, 6). Nucleotide sequencing of the mutant gene revealed an amino-acid substitution from the conserved alanine to threonine in the amino terminus (Supplementary Information Fig. 1c). The *cnd2-1* phenotype was recessive.

Immunoblotting indicated a single polypeptide (relative molecular mass of 105,000 (M_r 105K)) in the *Schizosaccharomyces pombe* cell extracts. The level of Cnd2 in the wild type was the same as in *cnd2-1* mutants incubated at 36 °C (Fig. 1a), indicating that the mutant phenotype was not caused by a decrease in the level of mutant protein.

To immunoprecipitate mutant protein, the *cnd2-1* gene was tagged at the carboxy terminus with eight times Myc or jellyfish green fluorescent protein (GFP), and chromosomally integrated. Using anti-Myc antibodies, immunoprecipitation of extracts prepared from cultures kept at 26 °C or at 36 °C for 4 h was performed for untagged Cnd2 (wild-type 972), tagged Cnd2-Myc⁶ and mutant Cnd2-1-Myc (Supplementary Information Fig. 1d). The precipitates from both wild-type (Cnd2-Myc) and mutant (Cnd2-1-Myc)

cells contained two other condensin components (Cnd3 and Cut3), as detected by polyclonal antibodies. The condensin complex thus was maintained in mutant cells even at 36 °C.

Moreover, sucrose gradient centrifugation (Fig. 1b) gave rise to a 13S peak for mutant extracts (incubated at 26 °C and 36 °C for 4 h), as in the wild type. Mutant Cnd2 protein sedimented together with other condensin subunits. Of note, Cnd2 was present mostly in the 13S fractions in both wild-type and mutant extracts. The mutant complex was unstable, as the 13S peak was significantly reduced in the extracts of *cnd2-1* cells cultured at 36 °C for 4 h. Hence, the principal fraction of Cnd2 was present in the condensin complex rather than as a free form or in other complexes.

In the wild type, Cdc2 promotes a marked mobilization of condensin from the cytoplasm to the nucleus at the point of entry into mitosis until telophase⁶. To determine whether mutant Cnd2 showed the same localization behaviour, GFP-tagged Cnd2-1 protein (produced by the chromosomally integrated gene) was visualized (Fig. 1c, lower panel). We observed accumulation of mutant

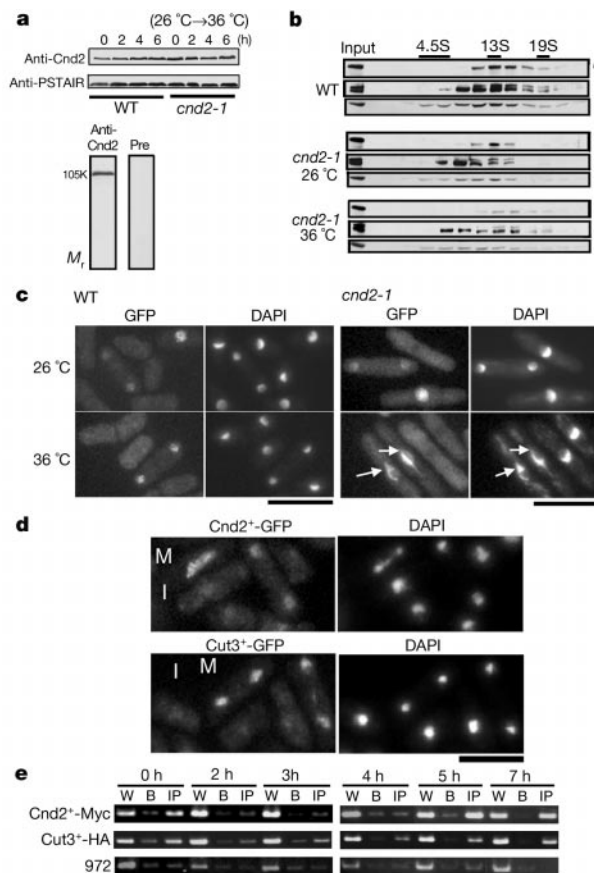


Figure 1 Isolation of *cnd2-1* and characterization of mutant protein. **a**, Polyclonal antibodies against GST–Cnd2 (full length) were used for detecting wild-type (WT) and mutant Cnd2 protein in cell extracts. Top, the level of Cnd2 protein in the cultures as assayed by immunoblotting; bottom, pre-immune serum (pre) does not produce the Cnd2 band. **b**, Sucrose gradient centrifugation of cell extracts from wild-type and mutant *cnd2-1* cells cultured at 26 °C and 36 °C for 4 h. **c**, C-terminal, GFP-tagged wild-type and mutant *cnd2* genes were integrated into the chromosome. Arrows indicate mutant Cnd2-1-GFP seen in mitotic nuclei. Scale bar, 10 μ m. **d**, Cnd2-GFP signals produced by the integrated gene (left) were detected in interphase cells (I) treated by Triton X-100¹⁹. Mitotic nuclear signals (M) of Cnd2-GFP were intense. The right panel shows DNA stained by DAPI in the same cells. **e**, ChIP analyses for tagged Cnd2-Myc and Cut3-HA in wild-type cells arrested at the G1 phase and released in complete medium. At 0 h (G1 cells under nitrogen starvation), a significant amount of Cnd2 and Cut3 seemed to be associated with Cnt1. FACS analysis indicates that the S and M phase occurs at 3–4 and 5 h, respectively. W, whole-cell extracts; B, beads; IP, immunoprecipitated; 972, untagged wild type.

* Present addresses: Department of Molecular and Cellular Biology, Howard Hughes Medical Institute, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA (T.S.); Department of Molecular Diagnosis (F8), Graduate School of Medicine, Chiba University, 1-8-1 Inohana, Chuo-ku, Chiba 260-0856, Japan (T.T.).

Cnd2-1-GFP protein in the mitotic nuclei (indicated by arrows), although mutant chromosomes neither condensed nor separated. Cnd2-1 protein thus behaved cytologically, as if it were a part of condensin, but failed to support chromosome condensation. (Wild-type (control) Cnd2-GFP localization is shown in the upper panel of Fig. 1c.)

In the nuclei of cells at interphase G2, the GFP signal of the condensin subunit was barely detectable⁶. However, if cells were treated with Triton X-100, which removed proteins unbound to chromatin¹⁶, interphase cells showed the weak nuclear signals of Cnd2-GFP and Cut3-GFP (indicated by I in Fig. 1d)—mitotic and

telophase signals were strong (indicated by M in Fig. 1d).

To confirm that both Cnd2 and Cut3 were bound to chromatin at interphase, chromatin immunoprecipitation¹⁷ (CHIP) was performed for interphase cell extracts using antibodies against Cut3-haemagglutinin (HA) and Cnd2-Myc, expressed by the tagged genes integrated on the chromosome (Fig. 1e; 972 was untagged wild-type control). Immunoprecipitation showed that both Cnd2-Myc and Cut3-HA derived from interphase chromatin (0–4 h after nitrogen-starved G1 cells were released to the complete medium) were co-precipitated with the probe DNA Cnt1. Cnt1 is a central centromere sequence that gave rise to the highest level of immunoprecipitates from mitotic cells among the many probes tested (data not shown). The intensity of immunoprecipitated Cnt1 DNA was significant in interphase (0–4 h; 3–4 h in the S phase), although less than that in mitotic cells (5–7 h).

Unexpectedly, *cnd2-1* was hypersensitive to ultraviolet irradiation (100 J m^{-2}), hydroxyurea (HU; 2 mM) and methylmethane sulphonate (0.003%) at 26 °C, which is the permissive temperature (Fig. 2a). Note that *cnd2-1* was highly sensitive to HU, comparable to the $\Delta rad3$ mutant, but its sensitivity to ultraviolet irradiation and methylmethane sulphonate was moderate. These

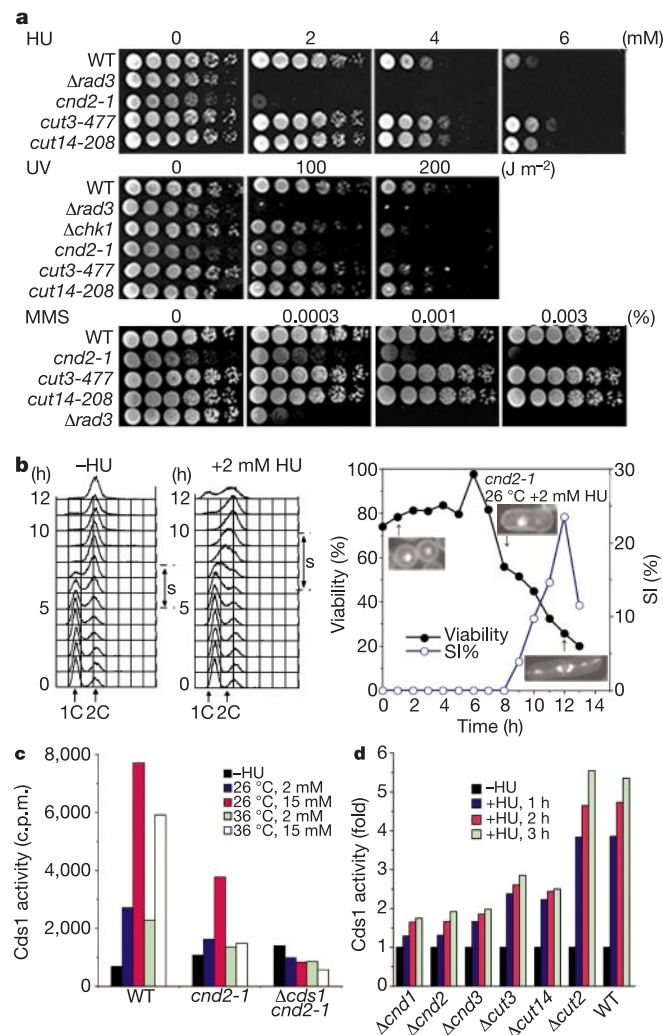


Figure 2 Cnd2 is required for the recovery from HU-induced arrest and the activation of Cds1 in the presence of HU. **a**, *cnd2-1* and other strains were spotted at 26 °C on the YPD medium containing 2–6 mM HU (top panel). The same strains were irradiated by ultraviolet light (middle panel) or spotted on YPD medium containing methylmethane sulphonate (MMS). **b**, *cnd2-1* cells were arrested at G1 at 26 °C by nitrogen starvation, and then released to the YPD medium at 26 °C in the absence (–HU) or presence (+HU) of 2 mM HU. Left, FACS analysis. Under the permissive condition (–HU, 26 °C), the S phase (S) occurs from 5 to 7 h. In the presence of HU (+HU, 26 °C), the S phase is delayed by 1 h (6–9 h), and coincides with the initial decrease of cell viability. **c**, Cds1 activity measured according to the procedure described³⁰. A peptide (RLARAASMAALARK) was made and used for the assay in extracts. The presence or absence of Chk1 did not affect the activity in extracts (data not shown). **d**, Spores derived from six heterozygous diploid strains containing gene disruptions of $\Delta cnd1$, $\Delta cnd2$, $\Delta cnd3$, $\Delta cut3$, $\Delta cut14$ or $\Delta cut2$ were cultured in the minimal medium, and spores with gene disruption were allowed to germinate. A concentration of 15 mM HU was added after 5 h for a period of 1–3 h. Extracts were prepared, and the Cds1 activity was measured.

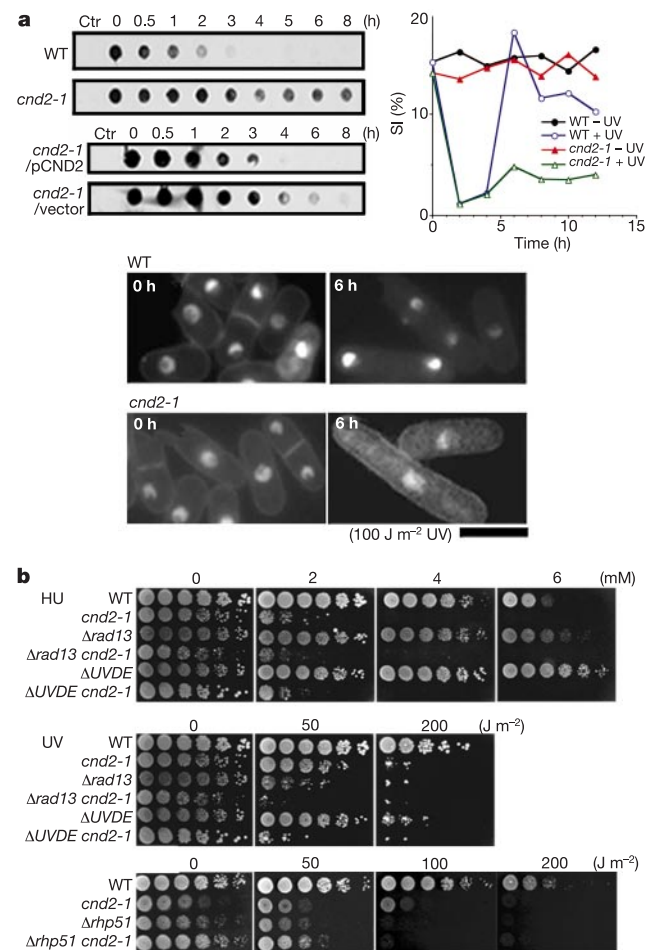


Figure 3 Cnd2 is required for repairing damaged DNA. **a**, Estimated levels of thymine dimers detected by antibodies (left). No thymine dimers were detected before ultraviolet (UV) irradiation (ctr), and levels were maximal immediately after irradiation (0 h) and disappeared after 3 h in wild-type (WT) cells. In *cnd2-1* cells irradiated at 26 °C, however, the level remained high even after 8 h. The defect was suppressed by plasmid pCND2. The SI% of *cnd2-1* at 26 °C after irradiation (green) sharply decreased and remained at a low level in *cnd2-1* (right). Mutant cells were elongated after irradiation (bottom panels). Scale bar, 10 μm . **b**, The sensitivity to ultraviolet irradiation was examined for single and double mutants of *cnd2-1* and deletion mutants of Rad13, UVDE and Rhp51.

sensitivities were restored by introducing plasmid pCND2 into *cnd2-1*. Condensin SMC mutants *cut3-477* and *cut14-208* were not sensitive to these agents.

To examine whether *cnd2-1* might reveal a severe S-phase defect in the presence of 2 mM HU, mutant cells were first arrested at G1 by nitrogen starvation at 26 °C, and then transferred to YPD medium in the presence or absence of 2 mM HU. Viability was lost during the S phase at 26 °C only in the presence of 2 mM HU (Fig. 2b). Aberrant mitosis lacking condensation was observed. Fluorescence-activated cell sorting (FACS) analysis indicated that the onset of S phase in the presence of 2 mM HU was delayed (by approximately 1 h), in comparison with that in the absence of 2 mM HU (left panel). In the wild-type control, no defects in viability or mitosis were found with 2 mM HU (data not shown). Cnd2 was thus essential for maintaining cell viability at 2 mM HU. At a much higher concentration of HU (15 mM), *cnd2-1* cells were arrested in the S phase, but cell viability decreased. In the wild type, cells did not lose viability at 15 mM HU (data not shown). To obtain precise information on the temporal progression of the cell cycle, a synchronous culture using an elutriator rotor was carried out: *cnd2-1* that was synchronized in G2 at 26 °C was released into 2 mM HU-containing YPD, and the cell-cycle progression was examined. The result was as expected. *cnd2-1* went through a normal first mitosis with normal timing, but the entry into the

second mitosis was delayed and resulting mitosis was abnormal (data not shown). The defect of *cnd2-1* in 2 mM HU at 26 °C was therefore produced after cells went through the S phase of the cell cycle once.

The checkpoint kinase Cds1 is markedly activated in HU-treated wild-type cells^{13,18}. To examine whether Cnd2 is implicated in this activation, we assayed the activity of Cds1 extracts in wild-type and *cnd2-1* cells (Fig. 2c; see legend for details). In the wild-type extracts, the activity greatly increased (10–12-fold) at 26 °C and 36 °C in the presence of 15 mM HU, but in the *cnd2-1* extracts, activation was significantly diminished at 26 °C and was hardly detectable at 36 °C. At 2 mM HU, activation was moderate in wild type, and diminished further in *cnd2-1*. Cnd2 is thus required for the activation of Cds1 in the presence of HU. We then measured the level of Mik1 kinase, which increased in the presence of HU in a Cds1-dependent manner¹⁹. The wild-type control showed an increase of Mik1, whereas in *cnd2-1* mutant cells the level did not increase at 36 °C in the presence of 15 mM HU (Supplementary Information Fig. 2). The reduced activation of Cds1 and the failure of Mik1 to increase in the presence of HU might explain why *cnd2-1* is sensitive to HU.

To address the question of whether other condensin subunits are also involved in these interphase phenotypes of *cnd2-1*, we used gene disruptions, because temperature-sensitive mutants were not available for other non-SMC subunits. Spores of six gene-disrupted

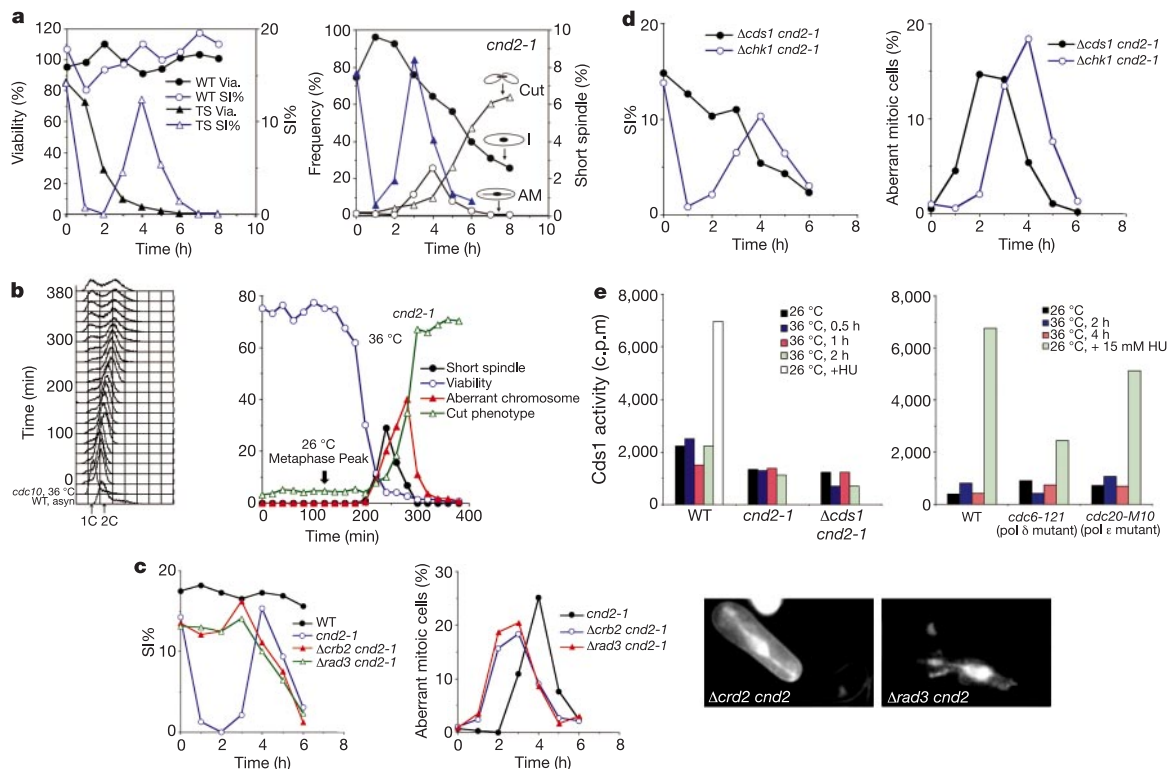


Figure 4 Interphase defect in *cnd2-1* at 36 °C monitored by Rad3-, Crb2- and Cds1-dependent checkpoint control. **a**, Mutant *cnd2-1* incubated at 26 °C was transferred to 36 °C. The left panel shows the SI% and cell viability (via.) of *cnd2-1* (open and filled triangles, respectively) at 36 °C and wild type control (open and filled circles, respectively). Nuclear chromatin stained by DAPI was classified into interphase nuclear chromatin (I), aberrant mitotic chromosomes (AM) and cut phenotypes (Cut) (right panel). The percentage of cells containing the short mitotic spindle is shown (filled triangles). **b**, Early G2 cells of *cnd2-1* incubated at 26 °C were collected by elutriation and used to perform synchronous culture at 36 °C. The left panel shows a FACS analysis indicating that the 2C DNA content decreases to 1C around 300 min in cells bisected by the cut phenotype. Cell viability sharply declines at 200 min, whereas aberrant mitosis occurs at 240–280 min followed by the cut phenotype (right panel). Asyn, asynchronous. **c**, The temporal arrest of

cnd2-1 at 36 °C is abolished in $\Delta rad3$ and $\Delta crb2$ mutant cells. The septation index (SI%) of double mutants does not decrease for the first 4 h after the shift to 36 °C (left panel). The decrease after 5 h is due to the occurrence of the cut phenotype. The peaks of aberrant mitosis in the double mutants at 36 °C are 2 h earlier than that in single mutant *cnd2-1* (middle panel). The right panel shows double mutant cells showing aberrant chromosomes stained by DAPI. **d**, The SI% and cytology were studied for the double mutants *cnd2-1* $\Delta chk1$ and *cnd2-1* $\Delta cds1$ cultured at 36 °C (see text for details). **e**, Cds1 activity was measured for extracts of *cnd2-1* cultured at 36 °C for 0–2 h in the absence of HU (left panel). Experiments were basically identical to that for Fig. 2c. The increase of Cds1 activity was not detected in extracts of DNA polymerase mutants at 36 °C (right panel). For a control, wild-type cells (WT) were cultured at 36 °C in the absence or the presence of 15 mM HU.

strains ($\Delta cut2$ is a control) were germinated, followed by culture in the presence of 15 mM HU for 0–3 h. We measured the activity of Cds1 in extracts (Fig. 2d). Wild-type and the $\Delta cut2$ control produced a 5–6-fold increase in the activity of Cds1, whereas deletions of the five condensin subunits produced 1.7–2.8-fold increases. These results were highly reproducible. Considering that a certain amount of Cds1 was brought into the spores from zygotic cells, the difference seems quite significant. Not only non-SMC subunits (Cnd1, Cnd2, Cnd3) but also SMC (Cut3, Cut14) subunits are required for the activation of Cds1, strongly suggesting that the whole condensin complex is involved in the activation of Cds1.

Next we found that DNA damage accumulated in $cnd2-1$ cells after ultraviolet irradiation at 26 °C (Fig. 3a, left). Antibody detection of thymine dimers generated by irradiation²⁰ revealed that the level of dimers remained high in $cnd2-1$ cells even at 8 h after irradiation. DNA damage checkpoints must be established in $cnd2-1$ cells at 26 °C, however, as irradiated cells were elongated (Fig. 3a, lower panel) and the septation index (an indicator for cells undergoing cell separation (SI%)) remained low (green line, right panel), probably owing to the failure to remove the damaged lesions.

The inability to repair the damage seems to be related to excision repair genes (Fig. 3b: Rad13 and Uve1 (UVDE) are nucleases for excision repair in fission yeast^{21,22}). The double mutants $cnd2-1 rad13$ and $cnd2-1 \Delta uve1$ were far more sensitive to ultraviolet irradiation than single mutants. The sensitivity to HU slightly increased in the double mutants, in comparison with the single mutant $cnd2-1$. The temperature-sensitive phenotype of $cnd2-1 \Delta rad13$ was shown at the decreased temperature (Supplementary Information Fig. 3a). These results indicate that Cnd2 and excision repair proteins might share an essential role, possibly functioning in a parallel way. Rhp51 is a Rad51 homologue that is involved in recombination repair²³. The double mutant $cnd2-1 \Delta rhp51$ did not show any additive change in the sensitivity to ultraviolet irradiation.

Rqh1 (identical to Hus2 and Rad12) is an ATP-dependent RecQ helicase similar to the Q helicase defective in human Werner and Bloom syndromes, and required for reversible S-phase arrest^{24,25}. We found that the double mutant $cnd2-1 \Delta rqh1$ was nearly synthetic-lethal at any temperature, forming minute colonies only at 26 °C (Supplementary Information Fig. 3b). Cnd2 and Rqh1 may share an essential function in the recovery from HU treatment. Cnd2-1 interacted with mutations defective in DNA replication: $cnd2-1$ was synthetically lethal to mutations in DNA ligase ($cdc17-K42$) and ribonucleotide reductase ($cdc22-c11$), both of which are essential for DNA synthesis (Supplementary Information Fig. 3c). Another mutant, $cdc6$ (DNA polymerase δ), was also synthetic-lethal with $cnd2-1$ at 33 °C (data not shown).

A surprising phenotype found for $cnd2-1$ cells initially grown at 26 °C and then shifted to 36 °C in rich medium was a 2-h delay of the cell cycle, the period of one generation. The septation index markedly decreased from 13% to less than 1% after the shift to 36 °C (Fig. 4a, left panel, open triangles). Cell viability (filled triangles) decreased sharply during the arrest. This temporal arrest requires the presence of $rad3$ and other DNA checkpoint genes (described below). Quantitative data of the cytological phenotypes (Fig. 4a, right panel) indicated that the frequency of the interphase cells decreased, whereas mutant cells entering aberrant mitosis (stained by DAPI) lacked condensation and contained the spindle (the short spindle index indicated by the filled triangles; stained by anti-tubulin antibody), followed by an increase in the cut phenotype (cell separation in the absence of normal nuclear division). Two features are notable: the temporal arrest of $cnd2-1$ was in interphase, and viability loss occurred during this temporal arrest.

We carried out synchronous culture analysis in order to investigate the phenotype in more detail (Fig. 4b). Early G2 cells were collected by elutriation from mutant cultures grown at 26 °C and transferred to 36 °C. The entry into mitosis was delayed by the

expected period (about 2 h). Occurrence of aberrant mitosis, in the absence of condensation, peaked at 280 min, whereas prometaphase and metaphase cells containing the short spindle peaked at 240 min. A sharp decline of viability occurred around 200 min, 40 min before metaphase. Control mutant synchronous cultures at the permissive temperature (26 °C) showed a normal metaphase peak (120 min; short thick arrow).

We next addressed the question as to how the delay was caused. A possibility was that a chromosomal problem occurs in $cnd2-1$ at 36 °C during interphase, resulting in the activation of checkpoint control and the prevention of entry into mitosis. Consistently, the delay at 36 °C was completely abolished when $rad3^+$ and $crb2^+$, the two DNA checkpoint genes in fission yeast^{10,20}, were deleted in $cnd2-1$. Rad3 is a homologue of ATM/ATR kinases, whereas BRCA1-like C-terminus (BRCT)-containing Crb2 interacts with Rad3 (refs 11, 20). The double mutants $cnd2-1 \Delta rad3$ and $cnd2-1 \Delta crb2$, constructed by crossing, were cultured at 36 °C. They showed aberrant mitosis without any delay at 36 °C (Fig. 4c). The septation index did not decrease in the double mutant cells (Fig. 4c, left panel), whereas aberrant mitosis could already be observed at 2 h—2 h earlier than that in single $cnd2-1$ mutants (middle panel). The aberrant chromosomes in the double mutants looked identical to those of the single mutants, although the timing of appearance was different (right panel). We thus concluded that the cell-cycle delay in $cnd2-1$ cells at 36 °C is due to a checkpoint control requiring the functions of Rad3 and Crb2. Cnd2 may be required for maintaining normal interphase chromosome structure, for example, by restoring spontaneously occurring defects in interphase.

Two protein kinases, Cds1 and Chk1, are implicated in inhibiting CDK kinase after the activation of checkpoint control^{12,13,18}. To determine whether $cnd2-1$ activated either or both Cds1- and Chk1-dependent checkpoints, the double mutants $cnd2-1 \Delta chk1$ and $cnd2-1 \Delta cds1$ were constructed. The temporal arrest of $cnd2-1$ at 36 °C remained in the double mutant $cnd2-1 \Delta chk1$, but was abolished in $cnd2-1 \Delta cds1$ (Fig. 4d). The timing of aberrant mitosis in $cnd2-1 \Delta cds1$ occurred 2 h earlier than that in $cnd2-1 \Delta chk1$ (right

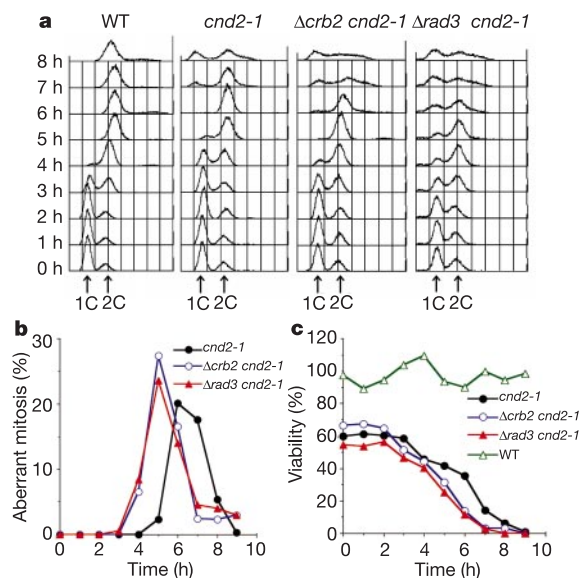


Figure 5 The role of Cnd2 in interphase. $cnd2-1$ cells arrested at G1 under nitrogen starvation at 26 °C were shifted to the YPD medium in the absence of HU at the restrictive temperature (36 °C). Wild-type control and two double mutants, $cnd2-1 \Delta crb2$ and $cnd2-1 \Delta rad3$, were also examined. **a**, FACS analysis showed that the S phase occurred at 3–4 h in wild type and the double mutants, but 4–5 h in the single $cnd2-1$ mutant. **b**, Quantitative data for the phenotypes of single $cnd2-1$ and the double mutant cells. **c**, The loss of cell viability at 36 °C was measured.

panel). We therefore concluded that the delay in *cnd2-1* at 36 °C was Cds1-dependent and not Chk1-dependent.

We then measured the activity of Cds1 in *cnd2-1* cells at 36 °C for 0–2 h (Fig. 4e). For a control, wild type and polymerase δ mutant *cdc6-121* and polymerase ϵ mutant *cdc20-M10* were used. These were cultured at 36 °C, and mutant cells were arrested after 2–4 h. The increase of the Cds1 activity could not be detected in *cnd2-1* and polymerase mutant extracts. Only a small activation of Cds1 might be sufficient for the arrest in the absence of HU, as the assay might not be sufficiently sensitive.

We next investigated whether the delay in *cnd2-1* could occur in the G1 phase at 36 °C. Nitrogen-starved G1 cells of *cnd2-1* at 26 °C were shifted to complete medium at 36 °C. FACS analysis showed that the entry into the S phase occurred at 4 h for *cnd2-1*, 1 h later than the wild type (Fig. 5a). The timing of the S-phase entry for the double mutants *cnd2-1 Δ rad3* and *cnd2-1 Δ crb2*, however, was 1 h earlier than that of the single *cnd2-1* mutant, suggesting that the 1 h delay in *cnd2-1* was dependent on these checkpoint genes. The timing for the appearance of aberrant mitosis and the decrease of cell viability was also 1 h earlier in the double mutants compared with that of the single *cnd2-1* mutant (Fig. 5b, c). The delay could thus take place in both G1 and G2 in a checkpoint-dependent manner.

This study demonstrates that Cnd2 is required for the process of DNA repair, the recovery from HU-induced S-phase arrest, and the activation of Cds1. In *cnd2-1* cells at 36 °C in the absence of HU, the Cds1- and Rad3-dependent checkpoint delay was induced. These interphase phenotypes together with the defect in mitotic condensation strongly suggest that Cnd2 has dual roles in both interphase and mitosis. Is this role restricted to Cnd2, one of the three non-SMC subunits? The lack of Cds1 activation in the deletion mutants of non-SMC and also SMC subunits showed that the whole complex is involved in the interphase functions. Why, then, do *cut3* and *cut14* mutants exhibit only mitotic defects? It was shown recently that the non-SMC subunits form stable heterotrimers that interact with the globular domains of SMC subunits in the heteropentameric condensin¹⁴. We propose that SMC mutations (*cut3-477* and *cut14-208*) that reside in the coiled-coil region²⁶ affect only mitotic condensation through mutual SMC interaction, whereas *cnd2-1* mutations affecting the interaction between non-SMC and the globular domains of SMC reduce the activity of condensin holocomplexes required for both interphase and mitotic functions. This hypothesis is well fitted to the model¹⁴ predicting that the ATPase domains of the SMC heterodimer is structurally regulated by the non-SMC trimer, and that the interaction between SMC heterodimers is crucial for mitotic condensation through formation of the large protein aggregates on DNA. The opening of the SMC heterodimer head may occur and be regulated by transient gate opening by means of the non-SMC trimer, in order to enclose the duplex DNAs for forming a loop¹⁴. Condensin in *cnd2-1* cells may fail in such reactions at 36 °C. Potential involvement of condensin SMC and non-SMC subunits in interphase nuclear functions was previously suggested in organisms such as the nematode, budding yeast and *Drosophila*^{9,27,28}, but no specific role has been reported to date.

Condensin complexes at interphase might function singly, but mitotic condensin complexes must interact with each other to form a large assembly on chromosomal DNA¹⁴. A large amount of condensin complex is thus mobilized into the nucleus on entry into mitosis⁶. Interphase condensin already present in the nucleus might recruit Cds1 and other unidentified proteins to appropriate sites in chromatin, so that cell viability is maintained after irradiation or in the presence of HU, or even after damages that are continuously generated in normal living cells. A relatively small amount of nuclear condensin might be sufficient for the interphase functions. □

Methods

Strains, plasmids, media and synchronous culture

Schizosaccharomyces pombe temperature-sensitive *cnd2-1* was isolated as described in the text. We verified that the mutation was in the *cnd2* gene as it was tightly linked to the *ura4⁺* marker integrated at the *cnd2* locus. The strains chromosomally integrated with the eight times Myc- or GFP-tagged mutant *cnd2* were generated in this study. Other strains and plasmids used were described previously^{6,20}. We used complete YPD and synthetic EMM2 media. Synchronous culture by elutriation and nitrogen starvation was done as described²⁹. The integrated *Mik1⁺*-HA strain and *Cds1* plasmids were gifts from P. Russell. Mutant strains *Δ rad13*, *Δ uve1*, *Δ rql1* and *Δ rhpl51* were gifts from A. M. Carr.

Microscopy, centrifugation, immunochemistry and CHIP

The procedures for fluorescence microscopy, sucrose gradient centrifugation and immunoprecipitation were described previously^{6,20}. Rabbit polyclonal antibodies were raised against recombinant glutathione S-transferase (GST)-tagged Cnd2 purified from bacterial cells. For detecting and observing mutant Cnd2-1 protein, the strains integrated with mutant Cnd2-1 tagged with eight times Myc and GFP were used, respectively. Methods to observe chromatin-bound protein by removing chromatin-unbound proteins by Triton X-100 were used¹⁶. Detailed procedures of the CHIP method were as described¹⁷.

HU sensitivity, ultraviolet irradiation and thymine dimer detection

The procedures for ultraviolet irradiation, HU sensitivity measurements and detection of thymine dimer were previously described²⁰.

Received 11 December 2001; accepted 27 February 2002.

- Hirano, T. & Mitchison, T. J. A heterodimeric coiled-coil protein required for mitotic chromosome condensation. *Cell* **79**, 449–458 (1994).
- Saka, Y. *et al.* Fission yeast cut3 and cut14, members of the ubiquitous protein family, are required for chromosome condensation and segregation in mitosis. *EMBO J.* **13**, 4938–4952 (1994).
- Strunnikov, A. V., Hogan, E. & Koshland, D. SMC2, a *Saccharomyces cerevisiae* gene essential for chromosome segregation and condensation defines a subgroup within the SMC-family. *Genes Dev.* **9**, 587–599 (1995).
- Kimura, K. & Hirano, T. ATP-dependent positive supercoiling of DNA by 13S condensin: a biochemical implication for chromosome condensation. *Cell* **90**, 625–634 (1997).
- Hirano, T. SMC-mediated chromosome mechanics: a conserved scheme from bacteria to vertebrates? *Genes Dev.* **13**, 11–19 (1999).
- Sutani, T. *et al.* Fission yeast condensin complex: essential roles of non-SMC subunits for condensation and Cdc2 phosphorylation of Cut3/SMC4. *Genes Dev.* **13**, 2271–2283 (1999).
- Bhat, M. A., Philp, A. V., Glover, D. M. & Bellen, H. J. Chromatid segregation at anaphase requires the barren product, a novel chromosome-associated protein that interacts with Topoisomerase II. *Cell* **87**, 1103–1114 (1996).
- Lavoie, B. D., Tuffo, K. M., Oh, S., Koshland, D. & Holm, C. Mitotic chromosome condensation requires Brn1p, the yeast homologue of Barren. *Mol. Biol. Cell* **11**, 1293–1304 (2000).
- Ouspenski, I., Cabello, O. A. & Brinkley, B. R. Chromosome condensation factor Brn1p is required for chromatid separation in mitosis. *Mol. Biol. Cell* **11**, 1305–1313 (2000).
- Bentley, N. J. *et al.* The *Schizosaccharomyces pombe* rad3 checkpoint gene. *EMBO J.* **15**, 6641–6651 (1996).
- Saka, Y., Esashi, F., Matsusaka, T., Mochida, S. & Yanagida, M. Damage and replication checkpoint control in fission yeast is ensured by interactions of Crb2, a protein with BRCT-motif, with Cut5 and Chk1. *Genes Dev.* **11**, 3387–3400 (1997).
- Murakami, H. & Okayama, H. A kinase from fission yeast responsible for blocking mitosis in S phase. *Nature* **374**, 817–819 (1995).
- Boddy, M. N., Furnari, B., Mondesert, O. & Russell, P. Replication checkpoint enforced by kinases Cds1 and Chk1. *Science* **280**, 909–912 (1998).
- Yoshimura, S. *et al.* Condensin architecture and interaction with DNA: regulatory non-SMC subunits bind to the head of SMC heterodimer. *Curr. Biol.* **12**, 508–513 (2002).
- Yamashita, Y. M., Nakaseko, Y., Kumada, K., Nakagawa, T. & Yanagida, M. Fission yeast APC/cyclosome subunits, Cut20/Apc4 and Cut23/Apc8, in regulating metaphase–anaphase progression and cellular stress responses. *Genes Cells* **4**, 445–463 (1999).
- Kearsey, S. E., Montgomery, S., Labib, K. & Lindner, K. Chromatin binding of the fission yeast replication factor mcm4 occurs during anaphase and requires ORC and cdc18. *EMBO J.* **19**, 1681–1690 (2000).
- Tomonaga, T. *et al.* Characterization of fission yeast cohesin: essential anaphase proteolysis of Rad21 phosphorylated in the S phase. *Genes Dev.* **14**, 2757–2770 (2000).
- Lindsay, H. D. *et al.* S-phase-specific activation of Cds1 kinase defines a subpathway of the checkpoint response in *Schizosaccharomyces pombe*. *Genes Dev.* **12**, 382–395 (1998).
- Christensen, P. U., Bentley, N. J., Martinho, R. G., Nielsen, O. & Carr, A. M. Mik1 levels accumulate in S phase and may mediate an intrinsic link between S phase and mitosis. *Proc. Natl Acad. Sci. USA* **97**, 2579–2584 (2000).
- Esashi, F. & Yanagida, M. Cdc2 phosphorylation of Crb2 is required for reestablishing cell cycle progression after the damage checkpoint. *Mol. Cell* **4**, 167–174 (1999).
- Carr, A. M. *et al.* Evolutionary conservation of excision repair in *Schizosaccharomyces pombe*: evidence for a family of sequences related to the *Saccharomyces cerevisiae* RAD2 gene. *Nucleic Acids Res.* **21**, 1345–1349 (1993).
- Yajima, H. *et al.* A eukaryotic gene encoding an endonuclease that specifically repairs DNA damaged by ultraviolet light. *EMBO J.* **14**, 2393–2399 (1995).
- Muris, D. F. *et al.* Cloning the RAD51 homologue of *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **21**, 4586–4591 (1993).
- Stewart, E., Chapman, C. R., Al-Khodairy, F., Carr, A. M. & Enoch, T. *rql1⁺*, a fission yeast gene related to the Bloom's and Werner's syndrome genes, is required for reversible S phase arrest. *EMBO J.* **16**, 2682–2692 (1997).

25. Murray, J. M., Lindsay, H. D., Munday, C. A. & Carr, A. M. Role of *Schizosaccharomyces pombe* RecQ homolog, recombination, and checkpoint genes in UV damage tolerance. *Mol. Cell Biol.* **17**, 6868–6875 (1997).
26. Sutani, T. & Yanagida, M. DNA renaturation activity of the SMC complex implicated in chromosome condensation. *Nature* **388**, 798–801 (1997).
27. Chuang, P. T., Albertson, D. G. & Meyer, B. J. DPY-27: a chromosome condensation protein homolog that regulates *C. elegans* dosage compensation through association with the X chromosome. *Cell* **79**, 459–474 (1994).
28. Lupo, R., Breiling, A., Bianchi, M. E. & Orlando, V. *Drosophila* chromosome condensation proteins Topoisomerase II and Barren colocalize with Polycomb and maintain Fab-7 PRE silencing. *Mol. Cell* **7**, 127–136 (2001).
29. Moreno, S., Hayles, J. & Nurse, P. Regulation of p34^{cdc2} protein kinase during mitosis. *Cell* **58**, 361–372 (1989).
30. Hutchins, J. R., Hughes, M. & Clarke, P. R. Substrate specificity determinants of the checkpoint protein kinase Chk1. *FEBS Lett.* **466**, 91–95 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank R. Yu for reading the manuscript. This study was supported by the CREST Research Grant of Japan Science and Technology Corporation (JST) and the COE Grant of the Ministry of Education, Culture, Science and Technology. T.S. and T.T. were the recipients of JSPS (Japan Science Promotion Society) and JST postdoctoral fellowships, respectively.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.Y. (e-mail: yanagida@kzo.biophys.kyoto-u.ac.jp).

Bacteriophytochrome controls photosystem synthesis in anoxygenic bacteria

Eric Giraud*, Joël Fardoux*, Nicolas Fourrier*, Laure Hannibal*, Bernard Genty†, Pierre Bouyer‡, Bernard Dreyfus* & André Verméglio‡

* LSTM TA 10/J Laboratoire des Symbioses Tropicales et Méditerranéennes, UMR 113, IRD/CIRAD/INRA/ENSA-M. Campus de Baillarguet, 34398 Montpellier Cedex 5, France

† CEA/Cadarache DEVM-Laboratoire d'Écophysiologie de la Photosynthèse, UMR 163-CNRS-CEA, 13108 Saint Paul lez Durance Cedex, France

‡ CEA/Cadarache DEVM-Laboratoire de Bioénergétique Cellulaire, UMR 163-CNRS-CEA, Univ-Méditerranée CEA1000, 13108 Saint Paul lez Durance Cedex, France

Plants use a set of light sensors to control their growth and development in response to changes in ambient light. In particular, phytochromes exert their regulatory activity by switching between a biologically inactive red-light-absorbing form (Pr) and an active far-red-light absorbing form (Pfr)^{1,2}. Recently, biochemical and genetic studies have demonstrated the occurrence of phytochrome-like proteins in photosynthetic and non-photosynthetic bacteria^{3–7}—but little is known about their functions. Here we report the discovery of a bacteriophytochrome located downstream from the photosynthesis gene cluster in a *Bradyrhizobium* strain symbiont of *Aeschynomene*. The synthesis of the complete photosynthetic apparatus is totally under the control of this bacteriophytochrome. A similar behaviour is observed for the closely related species *Rhodospseudomonas palustris*, but not for the more distant anoxygenic photosynthetic bacteria of the genus *Rhodobacter*, *Rubrivivax* or *Rhodospirillum*. Unlike other (bacterio)phytochromes, the carboxy-terminal domain of this bacteriophytochrome contains no histidine kinase features. This

suggests a light signalling pathway involving direct protein–protein interaction with no phosphorelay cascade. This specific mechanism of regulation may represent an important ecological adaptation to optimize the plant–bacteria interaction.

We have previously identified an unexpected haem oxygenase gene close to the photosynthesis gene cluster in the symbiotic photosynthetic *Bradyrhizobium* ORS278 strain⁸. The haem oxygenase opens the haem ring to form the linear tetrapyrrole biliverdin, the first intermediate in the synthesis of phytochrome chromophore⁹. This prompted us to search for the possible involvement of a bacteriophytochrome in the regulation of the synthesis of the photosynthetic apparatus in this species and other anoxygenic photosynthetic bacteria.

Sequencing the downstream region from the haem oxygenase gene of *Bradyrhizobium* ORS278 reveals, in the opposite orientation, an open reading frame (ORF) that encodes a polypeptide containing 724 amino acids (Fig. 1a). Its amino-terminal region displays a strong similarity to the chromophore-binding domain of

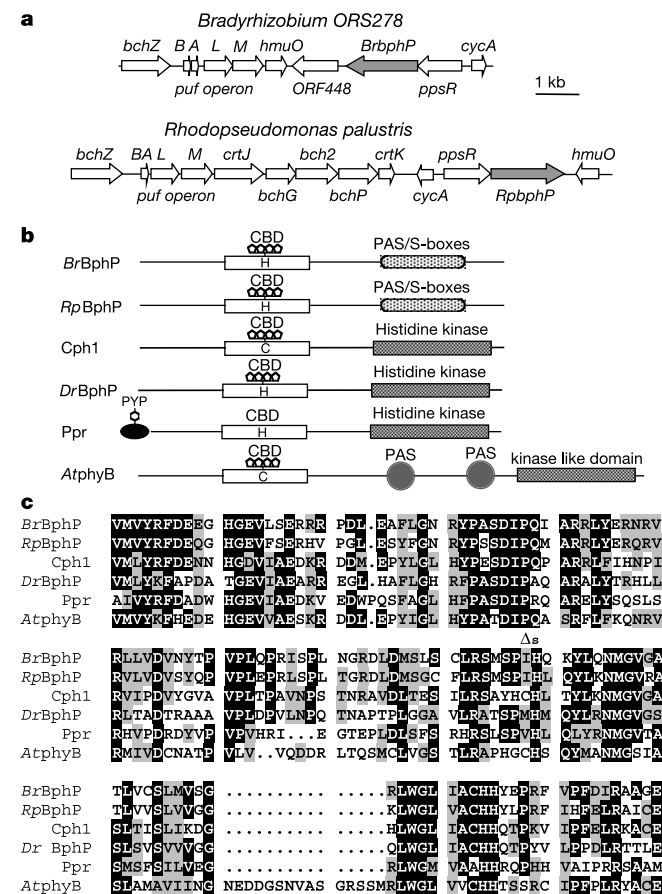


Figure 1 Molecular characterization of phytochromes. **a**, Arrangements of genes located downstream from the photosynthesis gene cluster in *Bradyrhizobium* ORS278 and *Rps. palustris*. *BrbphP*, *Bradyrhizobium* bacteriophytochrome photoreceptor; *RpbphP*, *Rps. palustris* bacteriophytochrome photoreceptor; *cycA*, gene encoding cytochrome *c*₂; *hmuO*, gene encoding haem oxygenase; *ppsR*, gene encoding the transcriptional factor PpsR. **b**, Domain structure of various phytochrome sequences. *BrBphP*, from *Bradyrhizobium* ORS278; *RpBphP*, from *Rps. palustris*; *Cph1*, from *Synechocystis* PCC6803; *DrBphP*, from *D. radiodurans*; *Ppr*, from *R. centenum*; and *AtphyB*, phytochrome B from *Arabidopsis thaliana*. CBD, chromophore binding domain; PAS, Per/Arnt/Sim repeats; PYP, photoactive yellow protein domain. **c**, Amino-acid sequence alignments of chromophore binding domains of *BrBphP*, *RpBphP*, *Cph1*, *DrBphP*, *Ppr* and *AtphyB*. The open and filled triangles indicate the conserved Cys and His residues involved in chromophore attachment in plant phytochromes and in *DrBphP*, respectively. Black background, identical residues. Grey background, similar residues.

bacterial and plant phytochromes (Fig. 1b), but lacks the conserved Cys residue expected to be the bilin ligand in plant phytochrome¹ (Fig. 1c). However, the adjacent His residue reported to serve as the bilin attachment site in *Deinococcus radiodurans* phytochrome (*DrBphP*) is conserved⁶. Unlike the case in other bacteriophytochromes, the C-terminal region of this protein, named *BrBphP* for *Bradyrhizobium* bacteriophytochrome photoreceptor, is not related to the histidine kinase domain, as none of the four conserved motifs that make up the catalytic centre is present¹⁰. However, an S-box domain that corresponds to a strongly conserved region in PAS (Per/Arnt/Sim) domains¹¹ is present from amino acid 518 to 642 (Fig. 1b). These domains, which may be involved in protein–protein interactions, are present in a large family of sensor proteins including the plant phytochromes¹². Immediately upstream of *BrbphP*, we identified an ORF encoding a 456-amino-acid protein homologous to the transcriptional factor PpsR (Fig. 1a). This protein represses the expression of bacteriochlorophyll (*bch*) and carotenoid (*crt*) genes and of light-harvesting II complex (*puc*) structural genes at high oxygen tension or high light intensity in anaerobic purple bacteria^{13,14}.

Purified recombinant *BrBphP* apoprotein was tested for *in vitro* attachment with various purified bilins by zinc-induced autofluorescence. The apoprotein covalently binds phycocyanobilin (PCB), phycoerythrobilin (PEB) and also biliverdin (BV), a bilin that cannot assemble with plant phytochromes¹⁵ but only with bacteriophytochromes¹⁶. The Pr form is produced in the dark immediately after addition of the chromophore, but is predominantly transformed into the Pfr form with a half-time of 30 min in a non-photochemical process (data not shown). Only the chromoproteins reconstituted with BV or PCB are photoactive and display reversible characteristic photoconversion between the Pr and Pfr forms (Fig. 2). We note that the Pfr form of the BV-bacteriophytochrome

presents the highest wavelength of maximum absorption ($\lambda_{\max} = 750$ nm) observed to date¹⁶ (Fig. 2). In agreement with the absence of a histidine kinase domain, no autophosphorylation could be observed in kinase assays *in vitro* for the active *BrBphP* using either red or far-red illumination.

To prove the functional involvement of *BrBphP* in the synthesis of the photosynthetic apparatus, *Bradyrhizobium* ORS278 cells were grown under continuous illumination with low irradiance of different wavelengths preferentially absorbed by the Pfr and Pr forms, or by the bacteriochlorophyll molecules (870 nm). Bacteriochlorophyll fluorescence was used as a specific marker of the presence of the photosynthetic apparatus (Fig. 3a). The wavelength dependence of photosystem synthesis observed *in vivo* (Fig. 3b) is characteristic of the absorption of the Pfr form of the BV-reconstituted *BrBphP* with an optimum close to 750 nm. This suggests that BV may be the chromophore of the native form of *BrBphP*. This experiment also indicates that Pr is the active form of this bacteriophytochrome. The enhancement of photosystem synthesis by the Pr form can be reversed to the level observed in the dark by applying 660-nm illumination at an irradiance approximately six times that of the 740-nm illumination, demonstrating the photoreversibility of the phytochrome effect. A further proof of the requirement of far-red illumination for the synthesis of photosynthetic apparatus was obtained by the precise and direct determination of the amount of light-harvesting complexes and reaction centres, by measuring the absorption spectrum and the light-induced photooxidation of

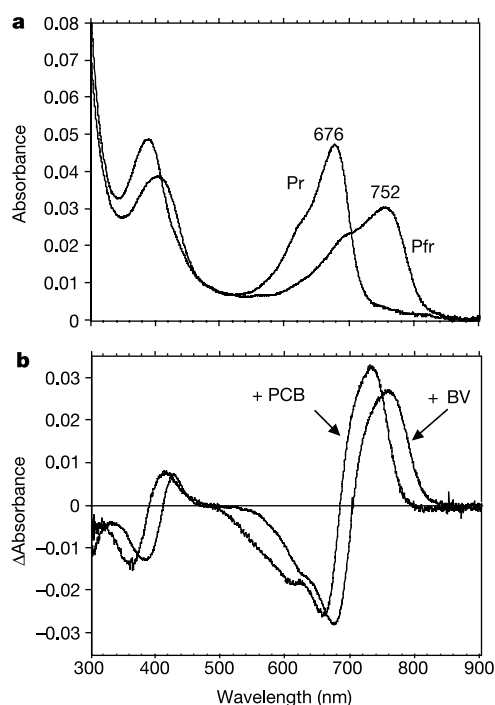


Figure 2 *In vitro* assembly of *Bradyrhizobium* bacteriophytochrome. **a**, Absorption spectra of Pr and Pfr forms of *Bradyrhizobium* bacteriophytochrome reconstituted with biliverdin (BV). The Pr spectrum was measured following far-red illumination ($\lambda > 740$ nm). The Pfr spectrum was derived from a spectrum recorded under 660-nm illumination assuming a Pfr/Pr ratio of 0.68. **b**, Difference spectra between the Pfr and Pr forms of *Bradyrhizobium* bacteriophytochrome reconstituted with BV or phycocyanobilin (PCB).

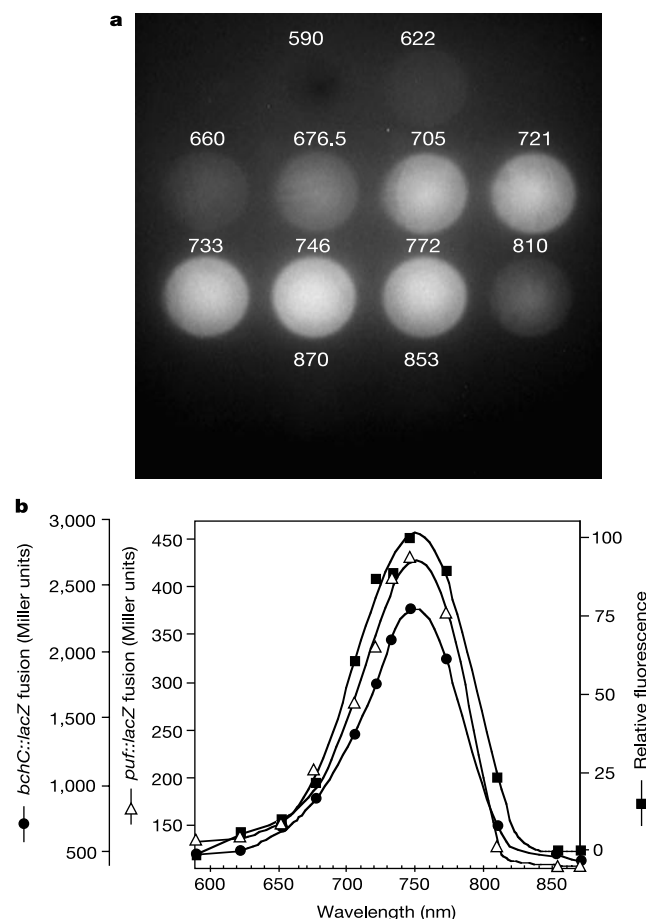


Figure 3 Action spectra for photosystem synthesis. **a**, Image of the long-wavelength fluorescence emission of bacteriochlorophyll of *Bradyrhizobium* cells grown on a Petri dish under light of different wavelengths (see Methods). **b**, Wavelength dependence of the synthesis of photosynthetic apparatus (filled squares) and expression of a *bchC::lacZ* (filled circles) and *puf::lacZ* (open triangles) fusions.

cytochrome, respectively, of intact cells (Fig. 4a). Furthermore, the expression of the *puf* operon and of the *bchC* gene—which encode the core proteins of the photosystem and an enzyme involved in bacteriochlorophyll synthesis, respectively—exhibited a wavelength dependence similar to the absorption spectrum of the Pfr form (Fig. 3b). This series of experiments shows that *BrBphP* controls the expression of various photosynthesis genes.

Proof of the involvement of *BrBphP* in the regulation of the photosystem synthesis by light was obtained with a null mutant of *BrBphP* generated by homologous-gene replacement. This mutant displays very low expression of the photosystem independent of the growth conditions (Fig. 4a). Altogether, these results give the molecular basis of the unexplained regulation of the photosynthetic apparatus synthesis by the light quality described for a symbiotic photosynthetic bacterium¹⁷.

To determine whether this system of regulation is widespread among anoxygenic photosynthetic bacteria, we analysed the behaviour of various purple bacteria belonging to the α or β subclass *vis-à-vis* the light quality. No significant effect of the light quality is observed on the synthesis of the photosynthetic apparatus of cells of *Rhodobacter* (*Rba.*) *sphaeroides*, *Rba. capsulatus*, *Rhodospirillum centenum* and *Rubrivivax gelatinosus*. In contrast, *Rhodopseudomonas* (*Rps.*) *palustris* exhibits a behaviour identical to that observed in *Bradyrhizobium* ORS278 strain. As an example, Fig. 4b shows a typical result for *Rps. palustris* and *Rba. capsulatus*. It is clear that the accumulation of the photosynthetic apparatus, which confers the red colour to the bacteria, requires illumination by 740-nm light for cells of *Rps. palustris* but not for *Rba. capsulatus*. The wavelength dependence of the expression of the photosynthetic apparatus of *Rps. palustris* is similar to that reported in Fig. 3 for *Bradyrhizobium* ORS278 (data not shown). This implies that the photosynthesis genes are also controlled by a bacteriophytochrome in *Rps. palustris*. In agreement with this, scanning the genomic database of *Rps. palustris* revealed the presence of a bacteriophytochrome gene homologous to *BrbphP* (Fig. 1b). This gene is adjacent to a *ppsR* gene and close to the photosynthesis gene cluster in an arrangement very similar to *Bradyrhizobium* (Fig. 1a). The presence of the gene coding for the transcriptional factor PpsR adjacent to the gene of the bacteriophytochrome in both photosynthetic bacteria (Fig. 1a) suggests that this protein could act as a regulator of the light signalling pathway. To test this hypothesis, we disrupted the *ppsR* gene of *Bradyrhizobium* ORS278. In contrast to the wild-type strain, the synthesis of the photosynthetic apparatus remained at a high level in the dark or under 660-nm illumination in a null *ppsR* mutant (Fig. 4a), demonstrating that PpsR blocks the transcription

of photosynthetic apparatus proteins in *Bradyrhizobium*. As the Pr form of *BrBphP* antagonizes the repressive activity of PpsR, we propose that these two proteins belong to the same light regulatory pathway, with *BrBphP* acting as the first and PpsR as the last element. Whether PpsR is the direct and only target of *BrBphP* has yet to be determined.

This series of experiments shows that *Rps. palustris* and photosynthetic *Bradyrhizobium* from *Aeschynomene* possess a sophisticated regulatory system where the synthesis of the entire photosynthetic apparatus is initiated by far-red light via the action of the Pr form of a bacteriophytochrome. This raises the question of why such a mechanism of regulation is only found in these two phylogenetically closely related species. A particular ecological niche and/or specific functional constraints may have favoured the establishment of this mechanism probably before their divergence. Such conditions are found in the case of photosynthetic *Bradyrhizobium* strains, which possess the exceptional capacity of nodulating both roots and stems of plants of the genus *Aeschynomene*¹⁷. In their stem symbiotic state, these bacteria grow below a layer of chlorophyll-containing cells, which absorbs preferentially blue and red light but transmits far-red light. Consequently, thanks to the control via a bacteriophytochrome, the light regulation of the photosystem synthesis allows the bacteria to switch rapidly from dark heterotrophy during root symbiosis to the more energetically favourable photoheterotrophy for stem symbiosis. As the bacterial photosynthetic activity is essential to the efficiency of the stem symbiosis⁸, this mechanism of regulation may represent an important ecological adaptation to optimize the plant–bacteria interaction. In the case of *Rps. palustris*—which possesses some *nod* genes but has not yet been shown to interact with plants—it remains to be determined whether the light regulation of its photosystem synthesis is linked to specific growth conditions or to the conservation of an ancestral character. □

Methods

Bacterial strains and growth conditions

Bradyrhizobium sp. strains were grown in a modified YM-agar medium⁸. *Rps. palustris*, *Rba. sphaeroides*, *Rubrivivax gelatinosus*, *Rba. capsulatus* and *Rhodospirillum centenum* were grown in Hutner medium¹⁸. All the strains were cultured in Petri dishes at 30°C in either complete darkness or continuous red (660 nm) or far-red (740 nm) light provided by light-emitting diodes (LEDs).

Wavelength dependence determinations

Bradyrhizobium cells, homogeneously inoculated on solid modified YM medium in a Petri dish, were grown for 6 d under continuous illumination provided by a series of LEDs of different peak wavelengths between 590 and 870 nm. Each LED illuminated a 3.5-cm² area. Irradiance and spectra were assessed using a spectroradiometer. Half-peak bandwidth was below 25 nm for all wavelengths. Irradiance was adjusted at 6.6 μ mol photons per m² per s to be below saturation for all tested wavelengths. The relative amounts of photosynthetic apparatus in the different illuminated areas were quantified by imaging bacteriochlorophyll fluorescence emission (above 800 nm) excited by blue-green illumination over the Petri dish using a cooled charge-coupled-device camera. The reporter gene *lacZ* was previously inserted into the *pufL* and *pufM* genes of ORS278⁸. For *lacZ::bchC* fusion, the *lacZ-kan^r* cassette¹⁹ was inserted into the unique *SalI* site of the *bchC* gene. Cells grown under the illuminated areas were resuspended in 3 ml of water, and β -galactosidase activity of *Bradyrhizobium* ORS278 strains harbouring the *bchC::lacZ* fusion or *puf::lacZ* fusion was measured as described²⁰.

Absorption and light-induced absorption change measurements

Light-induced absorption changes in intact cells or reconstituted *BrBphP*s were performed as previously described²¹. The amount of the photosynthetic apparatus in intact cells was determined by the measurement of the photooxidized cytochrome, the electron donor to the photosynthetic reaction centre, detected by absorbance change at 422 nm after a saturating flash. Absorption spectra of intact cells of the various species were recorded with a Cary 50 spectrophotometer.

Expression, purification and activation of *BrBphP*

BrbphP was amplified by polymerase chain reaction (PCR) from genomic DNA using the primers: 5'-AGGGGAGCCATATGCCCG TTCCGCTGACGACGCCA-3' and 5'-GCCGGCGGCTCTCC GCAACCTCCTCGTCTGCGAGCGATACCA-3'. PCR product was double digested with *NdeI* and *SapI* and ligated into the pTYBI vector (NEB) for expression in *Escherichia coli*. The *BrBphP*-intein fusion protein was purified using a chitin affinity column and self-cleaved by 30 mM dithiothreitol at 4°C. Holo-*BrBphP* was

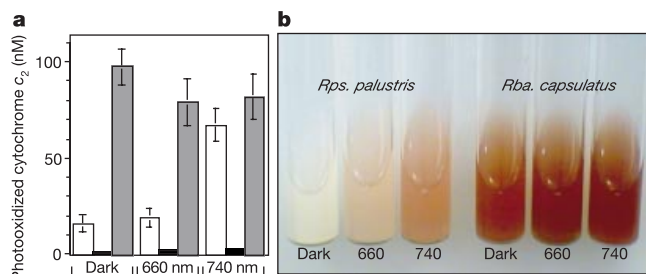


Figure 4 Effect of illumination on photosynthetic activity and pigmentation.

a, Photosynthetic activity of *Bradyrhizobium* ORS278 wild-type strain (unshaded), 278Δ*BrbphP* (black) and 278Δ*ppsR* (grey) strains grown in the dark, under 660-nm and 740-nm continuous illumination provided by LEDs. The data represent the mean of four experiments (errors bars indicate $\pm$ s.d.). **b**, Pigmentation of *Rps. palustris* and *Rba. capsulatus* cells after growth under complete darkness or continuous red light (660 nm) or far-red light (740 nm).

assembled with various bilins as described¹⁵. Spectra were obtained after saturating red (660 nm) and far-red (740 nm) irradiations. *In vitro* kinase assays for the active BrBphPs were performed as described⁵.

Construction of *BrbphP* and *ppsR* mutant strains

To create *Brbph* and *ppsR* null mutants, the *lacZ-kan^r* cassette¹⁹ was inserted respectively in the *XhoI* site of *BrbphP* and the *BglII* site of *ppsR*. The constructions were introduced in the pJQ200 suicide vector²² and delivered by conjugation into the ORS278 strain as previously described⁸. Double recombinants were selected on sucrose and confirmed by PCR.

Received 5 December 2001; accepted 19 February 2002.

1. Quail, P. H. *et al.* Phytochromes: photosensory perception and signal transduction. *Science* **268**, 675–680 (1995).
2. Smith, H. Phytochromes and light signal perception by plants—an emerging synthesis. *Nature* **407**, 585–591 (2000).
3. Kehoe, D. M. & Grossman, A. R. Similarity of a chromatic adaptation sensor to phytochrome and ethylene receptors. *Science* **273**, 1409–1412 (1996).
4. Hughes, J. *et al.* A prokaryotic phytochrome. *Nature* **386**, 663 (1997).
5. Jiang, Z.-Y. *et al.* Bacterial photoreceptor with similarity to photoactive yellow protein and plant phytochromes. *Science* **285**, 406–409 (1999).
6. Davis, S. J., Vener, A. V. & Vierstra, R. D. Bacteriophytochromes: phytochrome-like photoreceptors from nonphotosynthetic eubacteria. *Science* **286**, 2517–2520 (1999).
7. Schmitz, O., Katayama, M., Williams, S. B., Kondo, T. & Golden, S. S. CikA, a bacteriophytochrome that resets the cyanobacterial circadian clock. *Science* **289**, 765–768 (2000).
8. Giraud, E., Hannibal, L., Fardoux, J., Verméglio, A. & Dreyfus, B. Effect of *Bradyrhizobium* photosynthesis on stem nodulation of *Aeschynomene sensitiva*. *Proc. Natl Acad. Sci. USA* **97**, 14795–14800 (2000).
9. Muratomo, T., Kohchi, T., Yokota, A., Hwang, I. & Goodman, H. M. The *Arabidopsis* photomorphogenesis mutant *hyl* is deficient in phytochrome chromophore biosynthesis as a result of a mutation in a plastid heme oxygenase. *Plant Cell* **11**, 335–347 (1999).
10. Parkinson, J. S. & Kofoed, E. C. Communications modules in bacterial signalling proteins. *Annu. Rev. Genet.* **26**, 71–112 (1992).
11. Zhulin, I. B., Taylor, B. L. & Dixon, R. PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox. *Trends Biochem. Sci.* **22**, 331–337 (1997).
12. Taylor, B. L. & Zhulin, I. B. PAS domains: Internal sensors of oxygen, redox potential, and light. *Microbiol. Mol. Biol. Rev.* **63**, 479–506 (1999).
13. Ponnampalam, S. N., Buggy, J. J. & Bauer, C. E. Characterization of an aerobic repressor that coordinately regulates bacteriochlorophyll, carotenoid, and light harvesting-II expression in *Rhodobacter capsulatus*. *J. Bacteriol.* **177**, 2990–2997 (1995).
14. Gomelsky, M. & Kaplan, S. Molecular genetic analysis suggesting interaction between AppA and PpsR in regulation of photosynthesis gene expression in *Rhodobacter sphaeroides*. *J. Bacteriol.* **179**, 128–134 (1997).
15. Li, L. & Lagarias, J. C. Phytochrome assembly. *J. Biol. Chem.* **267**, 19204–19210 (1992).
16. Bhoo, S.-H., Davis, S. J., Walker, J., Karniol, B. & Vierstra, R. D. Bacteriophytochromes are photochromic histidine kinases using a biliverdin chromophore. *Nature* **414**, 776–779 (2001).
17. Fleischman, D. & Kramer, D. Photosynthetic rhizobia. *Biochim. Biophys. Acta* **1364**, 17–36 (1998).
18. Clayton, R. K. The induced synthesis of catalase in *Rhodospseudomonas sphaeroides*. *Biochim. Biophys. Acta* **37**, 503–512 (1960).
19. Kokotek, W. & Lotz, W. Construction of a *lacZ*-kanamycin-resistance cassette, useful for site-directed mutagenesis and as a promoter probe. *Gene* **84**, 467–471 (1989).
20. Boivin, C., Camut, S., Malpica, C. A., Truchet, G. & Rosenberg, C. *Rhizobium meliloti* genes encoding catabolism of trigonelline are induced under symbiotic conditions. *Plant Cell* **2**, 1157–1170 (1990).
21. Yurkov, V., Schoepp, B. & Verméglio, A. Photoinduced electron transfer and cytochrome content in obligate aerobic phototrophic bacteria from genera *Erythromicrobium*, *Sandaracinobacter*, *Erythromonas*, *Roseococcus* and *Erythrobracter*. *Photosynth. Res.* **57**, 117–128 (1998).
22. Quandt, J. & Hynes, M. F. Versatile suicide vectors which allow direct selection for gene replacement in gram-negative bacteria. *Gene* **127**, 15–21 (1993).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.V. (e-mail: avermeglio@cea.fr). GenBank accession codes for the (bacterio)phytochrome sequences are AF182374 (*Bradyrhizobium* ORS278), AB00139 (*Synechocystis* PCC6803), AAF12261 (*D. radiodurans*), AF064527 (*R. centenum*), X17342 (*Arabidopsis thaliana*). The genomic

organization of *Rps. palustris* was deduced from the genome database at <http://spider.jgi-psf.org/JGI-microbial/html/>.

corrigendum

Nuclear translocation and transcription regulation by the membrane-associated guanylate kinase CASK/LIN-2

Yi-Ping Hsueh, Ting-Fang Wang, Fu-Chia Yang & Morgan Sheng

Nature **404**, 298–302 (2000).

In this Letter, we numbered some nucleotides for the upstream region of the *reelin* gene incorrectly. The Reelin-luc construct contains an upstream region of the *reelin* gene corresponding to nucleotides 157700–158620 of human BAC clone AC002067, instead of nucleotides 3700–4620. This does not affect any of the results or conclusions of the paper. We thank A. M. Goffinet, D. Grayson, K. Mendra and T. Curran for alerting us to this mistake. □

erratum

Origins and estimates of uncertainty in predictions of twenty-first century temperature rise

Peter A. Stott & J. A. Kettleborough

Nature **416**, 723–726 (2002).

On page 725 of this Letter, the words ‘predicts¹³ fThur lglk al seasitevier’ were corrupted. They should read ‘predicts¹³. This lack of sensitivity’. □

Separate cells

From antibodies and assays to simple separations.

ProTeam FFE

Tecan www.tecan.com

Reduce the complexity of the proteome

ProTeam FFE is a liquid fractionation device that allows rapid separation of a wide range of charged molecules and particles without the use of a conventional solid polyacrylamide matrix. This technology may be used to separate cells, cell organelles, cellular fragments and complex protein mixtures. ProTeam FFE permits fast continuous electrophoretic fractionation yielding up to 96 fractions. Experiments can be performed with different electrophoretic modes under both native and denaturing conditions. ProTeam FFE is supplied with proprietary reagents and validated protocols for isoelectric focusing. Further protocols for zone electrophoresis and isoelectrophoresis operation modes are possible.

EBP50

Affinity BioReagents www.bioreagents.com

A key protein in cell signalling

Ezrin-radixin-moesin binding phosphoprotein of 50 kDa (EBP50) is a PDZ-containing protein that is involved in the linkage of integral membrane proteins to the cytoskeleton. This key protein is implicated in the formation of a wide variety of signalling complexes. EBP50 can be used in a range of cell signalling, differentiation and growth models. It has been optimized for immunoprecipitation and western blot procedures.

Haemoglobin antibodies

Biodesign www.biodesign.com

Antibodies to fetal and human haemoglobin

Biodesign's fetal haemoglobin (HbF) monoclonal antibodies detect the human fetal haemoglobin molecule in its intact form. Reactivity is demonstrated with human fetal erythrocytes, adult F cells and nucleated red cell precursors containing HbF. No reaction is observed with other forms of human haemoglobin. The binding epitope is present on the intact HbF molecule only. Human haemoglobin monoclonal antibodies detect many types of human haemoglobin, including HbA, HbA1c, HbA2, HbE, HbF and HbS.

Lysozyme

Pierce www.piercenet.com

Breakdown assistance

Lysozyme can be used with Pierce's B-PER bacterial protein extraction products to aid in the breaking down of cell walls and membranes in sample preparations and to help



Matrix-free fractionation from Tecan

clean inclusion body preparations. Lysozyme is produced from chicken egg white and has a specific activity of ~20,000 units per mg. One unit of enzyme catalyses a decrease of 0.001 A_{450} per cm per minute at 25 °C and pH 6.24 due to lysis using a suspension of *Micrococcus luteus* (about 0.25 mg per ml) as substrate.

Anti-TRAIL –1 to –4

Alexis Biochemicals www.alexis-corp.com

Receptor flow cytometry set

This monoclonal antibody set is used in the study of TNF-related apoptosis-inducing ligand (TRAIL) mediated apoptosis. These receptor-specific antibodies to TRAIL–R1, –R2, –R3 and –R4 can be used in the analysis of cell lysates by western blotting and cell surface studies using flow cytometry (FACS). While detection of TRAIL receptors in western blots can be performed with polyclonal sera, FACS applications appear to require MAbs specific to each TRAIL receptor. The MAbs for TRAIL–R1 and –R2 can also be used to identify

which of the two pro-apoptotic pathways are activated by TRAIL in a particular cell type.

Guava PC

Genetic Research Instrumentation

www.gri.co.uk

Up close and personal

Guava's PC personal cytometer is now available from Genetic Research Instrumentation. This flexible desktop cytometer can be used together with validated reagents and dedicated software to perform advanced single-cell analysis on a micro-volume scale. Applications range from cell counting and viability to protein expression and apoptosis analysis.

AutoLead

Imaging Research www.imagingresearch.com

Give cell analysis a boost

The AutoLead cell analyser is designed to increase the breadth and throughput of cell-based functional assays. It is supplied with bright field and epifluorescence optics and a rapid, high-resolution CCD camera. Software is available for common assays. The analyser processes over 2,500 samples per hour in 384-well microplates.

IMAP

Molecular Devices www.moleculardevices.com

Versatile mix-and-read assay kits

Molecular Devices offers three assay kits using IMAP technology, an assay format for kinase activity that does not use specific antibodies. IMAP can be applied to most kinases as a non-radioactive, mix-and-read assay. Kits include SGK and Akt assays for screening kinase inhibitors, and IMAP Explorer, which provides the company's assay tools.

These notes are compiled in the Nature office from information provided by the manufacturers.

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Laura Pearson (4977)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel + 33 (0) 1 43 20 16 51
Fax + 33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mails: p.phelan@nature.com
k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatomachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.cowan@naturejp.com
Japan Manager: Kate Cowan

naturejobs

An industrial revolution

Earning a prestigious graduate research fellowship from the US National Science Foundation (NSF) was once considered to be a ticket to an academic career. But a recent NSF report shows that the recipients are increasingly having second thoughts about capitalizing on their perceived advantage.

Almost all of the students surveyed thought when they received the award that they would pursue tenure-track positions at universities. But almost a quarter of the fellows who answered an open-ended question about career aspirations said they had changed their mind.

Even though the respondents came from a variety of disciplines — including engineering, economics and chemistry — they consistently gave three reasons for this switch. First, academia offers too few jobs. One reported initially being “behind a veil of ignorance” about how competitive the academic market was. Others complained about the low pay. “I’d be lying if I said that money wasn’t an issue” for considering industry, one respondent said. And several remarked that when they started their studies they were unaware of opportunities in government or business.

The fellowship’s biggest problem — that the stipend of \$15,000 is too low — is perhaps the most easily rectified, if the US government has the money and the will to spend it. But the overriding issue of the ‘best and the brightest’ being siphoned out of the academic pipeline as a result of disillusionment is tougher.

As one professor interviewed as part of a follow-up to the survey said, students are more open to options outside academia. If universities want to attract the top talent, they will have to stop taking them for granted.

Paul Smaglik

Naturejobs editor



Contents

SPECIAL REPORT

Making the most of the
National Institutes of Health p4

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

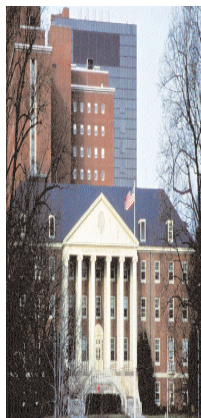
FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Every year, thousands of researchers head for the US National Institutes of Health. But what draws them there? Karen Kreeger finds out.

The US National Institutes of Health (NIH) in Bethesda, Maryland, at the start of the twenty-first century could be considered the biomedical equivalent of Ellis Island at the turn of the twentieth — a staging ground for people seeking their fortune in the New World. In 2000 alone, 2,500 researchers from 90 countries spent time there as visiting fellows.

This group mirrors the aspirations of the international postdoc community at large. Postdocs generally come to the NIH to learn new techniques, gain access to better research opportunities, or to work with a particular mentor.

Each individual's choice to come to Bethesda rests on a varied mix of those criteria, as well as a host of personal choices. But the decision also reflects the conditions in their home country, such as a lack of opportunities or resources. And those conditions make their next decision — looking for a foothold in the North American job market or finding a permanent position back home — that much more crucial.

CLASSIC PATH

Jacqueline Wood came from Britain to the NIH because she heeded the classic advice of picking the best lab and mentor to suit her interests. “I looked to see who was doing what I wanted to do,” Wood says. She found Jordan Grafman at the National Institute of Neurological Disorders and Stroke.

It helped that the NIH had equipment that allowed her to pursue her interest in using functional magnetic resonance imaging to study information storage in the brain's prefrontal cortex and its effects on behaviour. She says that getting a postdoc position in Britain that provides access to such equipment is difficult — especially if you have not done much imaging work before and if your work is not directly clinical.

After her three-year NIH postdoc, Wood must return to Britain because of the provisions of her visa. She says that there are reasons

both for wanting to return home and for trying to stay on in the United States. She cites the greater number of funding opportunities in the United States as a big draw.

There is also “less of a glass ceiling for women in science here in terms of research and salary”, she says. “Also I’m married, and the impression I get is that it’s much easier to solve the ‘two-body problem’ here.” Wood’s husband, Peter Brasted, is a postdoc at the National Institute of Mental Health.

Family reasons would take her back to the United Kingdom, she says, although jobs and funding might not be as easy to find as in the United States. But Britain has a strong research base in cognitive psychology, so jobs and funding in this field may improve, she thinks — but perhaps not enough to keep her there. “I certainly would go back, and then look to return to the United States because of the facilities and funding,” Wood says.

POINT OF NO RETURN

Herminia Gonzalez, a Spanish postdoc at the National Heart, Lung, and Blood Institute, had similar reasons for seeking out an NIH postdoc. But she has different prospects for returning home after she finishes working in the labs of Bryan Brewer and Silvia Santamarina-Fojo in the Molecular Disease Branch on animal models of cardiovascular diseases and heart attacks.

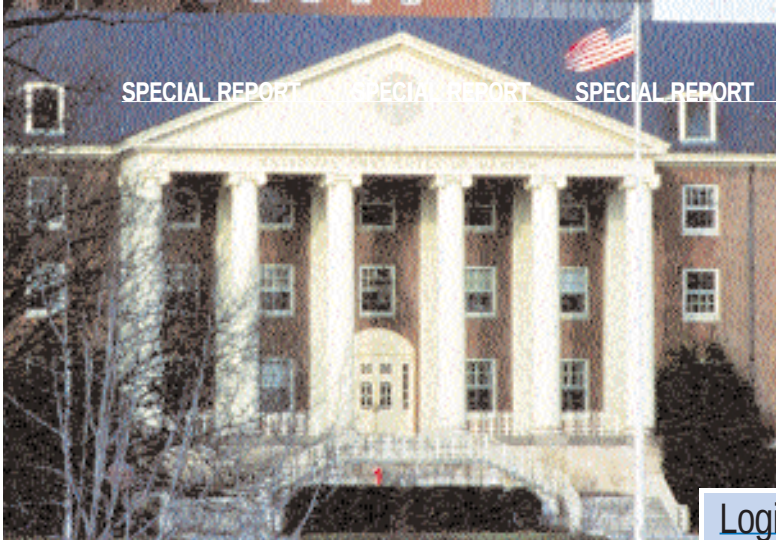
Gonzalez would definitely like to return to Spain, but says that is a “difficult thing”. There are not many possibilities for a permanent research position in Spain, she says. In the past, one possibility was to teach at a public university and do some part-time research, but this avenue has all but dried up. The other prospect she mentions is to work at public institutions, but jobs there are also “very limited and competitive”, says Gonzalez.

To alleviate this bottleneck, the Spanish government last year created ‘Ramon y Cajal’ five-year contracts (500 this year) to be offered to PhDs, whether Spanish natives or not, in all of the physical and biological sciences (see *Nature* **410**, 1014; 2001). Scientists who win a contract work within established groups at universities or public research ministries.

But even these positions don’t guarantee security. “After the five years nobody knows if the candidate will have a job, and it will depend on the number of

Matchmaker: Sharon Hrynkow (left) tries to target visiting fellows.





Logistical assistance

publications," says Gonzalez. The job prospects for life scientists in the private sector in Spain are not much better, in contrast to the United States and elsewhere in Europe, she says.

HOMEWARD BOUND

Andrea Baccarelli, an Italian postdoc in the genetic epidemiology branch of the National Cancer Institute (NCI), feels good about his chances of returning. "My main objective is to go back to Milan, where I was working before coming here," he says.

This is Baccarelli's first postdoc, but he already has his MD, which he received from the University of Perugia in 1995. He worked in Italy as a physician for five years in endocrinology before enrolling in a PhD programme in environmental and occupational medicine at the University of Milan.

His postdoc at the NIH (so designated because he already has an MD) is part of his PhD work assessing the effects of dioxin in humans. After completing his PhD, Baccarelli expects to work in epidemiology, as he is following a cohort of people who were exposed to dioxin after an industrial accident at Seveso in Italy in 1976.

He switched to the NCI because of a collaboration in molecular epidemiology with the Research Center for Occupational, Clinical and Environmental Epidemiology at the University of Milan. He is one year into his two-year NIH contract. "I could extend it, but I also need to go back to finish my PhD programme in Italy," he says. "I feel that I'm creating my own expertise being here and gaining a deeper knowledge of genetics, toxicology and epidemiology. This expertise, I think, will be needed in my department in Italy, and I'm pretty confident that I'll be able to run my own research there."

BIOTECH, BUT WHERE?

Many of the same reasons that attract other European postdocs to the United States — well-equipped labs and a mentor with complementary interests — also appealed to Detlef Vullhorst, a German postdoc at the National Institute of Child Health and Human Development. But different career goals might lead him to stay.

He is not pursuing an academic position back home, at least partly because even though the German *Habilitation* system — which requires a post-PhD thesis

Once a postdoc has landed a coveted fellowship at the US National Institutes of Health (NIH), the scientific tasks of setting up new experiments may seem like a cakewalk compared with the more mundane tasks of finding accommodation, living on a postdoc salary, and negotiating the intricacies of visa rules.

The NIH's John E. Fogarty International Center helps to facilitate the stay of visiting fellows "to the extent that we learn of highly qualified individuals who would like to make a match in the Visiting Program," says Fogarty's deputy director, Sharon Hrynkow, although lab chiefs make the "final call".

The centre works closely with the NIH

Office of Intramural Research to act as a broker, she adds. It also sends a handbook to prospective postdocs before they even arrive in the United States that provides information on the cost of living, how to find housing, visas, airports and taxes.

As far as making ends meet, intramural NIH stipends for US and foreign postdocs are identical. Starting in May 2002 the range will be \$33,000–55,200, and may be higher for physicians with additional clinical training or for certain areas of speciality. Moving expenses are generally not paid, although the NIH hosts a website that lists available accommodation

K.K.

The John E. Fogarty International Center
♦ www.nih.gov/fic

Jacqueline Wood (right) went to the NIH to work with Jordan Grafman.

— is being phased out, its replacement may be just as cumbersome (see *Nature* **415**, 257–258; 2002). If he does go back, he suspects it will be in biotechnology.

But Vullhorst, who has been at the NIH for just over two years, feels that the biotech climate is more favourable in the United States than it is in Germany, although it is improving there. "Germany is now at least next to Great Britain in terms of being the most active in biotechnology in Western Europe," he says.

For now, Vullhorst, like postdocs from anywhere, is mainly concerned with the task at hand. "I have not really educated myself and updated my knowledge about jobs in Europe as much as I should have," he says. "Right now I have to get to my papers." And, in doing so, hope that the future works itself out.

Karen Kreeger is a Philadelphia-based freelance science writer.

Detlef Vullhorst might stay in the United States.

